

# A Novel Chamfer Template Matching Method Using Variational Mean Field

Duc Thanh Nguyen  
Faculty of Information Technology  
Nong Lam University, Ho Chi Minh City, Vietnam

thanhnd@hcmuaf.edu.vn

## Abstract

*This paper proposes a novel mean field-based Chamfer template matching method. In our method, each template is represented as a field model and matching a template with an input image is formulated as estimation of a maximum of posteriori in the field model. Variational approach is then adopted to approximate the estimation. The proposed method was applied for two different variants of Chamfer template matching and evaluated through the task of object detection. Experimental results on benchmark datasets including ETHZShapeClass and INRIAHorse have shown that the proposed method could significantly improve the accuracy of template matching while not sacrificing much of the efficiency. Comparisons with other recent template matching algorithms have also shown the robustness of the proposed method.*

## 1. Introduction

Chamfer template matching is a well-known technique often used in many computer vision tasks, e.g. object detection [10] and recognition [23]. This is due to the simplicity and efficiency of the method. In addition, compared with learning based methods, e.g. [7], object detection using Chamfer template matching is often preferred in applications where the detection is required to perform using a single template supplied by the user and off-line learning every possible object class is impossible. Moreover, the template is unknown beforehand by the detection system.

Conventionally, contour templates are used to represent the object of interest and matching a template with an image can be performed through the distance transform (DT) calculating the spatial distance between template points and edge pixels on the input image [6].

A well-known challenge of template matching is the variation of the object shape which cannot be fully represented by templates. In addition, due to the sensitivity of edge detectors, e.g. Canny's detector [3], to illumination conditions and cluttered images, important edges of the ob-

ject shape may be missed while noisy edges from cluttered backgrounds may be presented. To overcome these difficulties, advanced developments of template matching have been proposed.

To cope with the local deformation of the object shape, Bai et al. [1] proposed the use of "shape band", a dilated version of templates corresponding to various deformed shapes of the object. However, the shape band does not constraint the location of template points on the same shape. In [23], shape context [2] and the continuity of object shape were used in template matching.

Attempts in improving the accuracy of template matching in cluttered images have been the use of edge orientation in complement with the spatial information. For example, in [9, 21], edge orientation was quantised and the DT was then computed for each quantised orientation. However, calculating the DT for every discrete orientation increases the computational cost. To relax the computational burden of such an operator, in [22, 19], the DT was used to find spatially nearest edge points of the given template and the orientation of those edge points was augmented with the spatial distance in computing the matching score. To obtain further improvement, in [19], edge magnitude was employed to weight edge points during calculating the DT. In [16], similarly to [21], three-dimensional DT computed over the location and orientation of edge pixels was employed. However, the three-dimensional DT was computed jointly in both the spatial and orientation domain using dynamic programming with integral images. In [17], false alarms were removed by matching the input image with random templates, i.e. templates not representing the object of interest.

Although the use of edge orientation with DT could enhance the accuracy of template matching, there are still remained issues. Essentially, matching a template with an edge image using the DT is to search for a set of edge points which are spatially close to template points. However, template points are matched independently. Thus, the best matching edge points obtained using the DT are not necessary to form any regular object in comparison to the

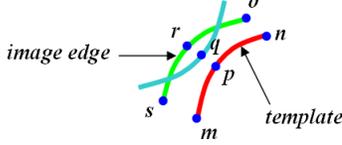


Figure 1. Illustration of a false matching by using only the distance transform. This figure is best viewed in colour.

given template. This problem becomes more challenging in cluttered images. Figure 1 illustrates a false matching case in which  $q$  would be considered as the best matching point of  $p$  using the DT. However,  $r$  actually represents a better match when the set  $\{s, r, o\}$  is compared with the template part  $\{m, p, n\}$ .

To overcome the above problem, we allow each template point/line<sup>1</sup> to have more than one matching candidates (i.e. closest edge points on the input image). The best matching candidate edge point/line on the image will be selected so as it is close to the template point/line and at the same time does not much deform the local shape formed by it in comparison to the template. This means that an edge point/line is not selected based only on its distance to the template but on its neighbouring edge points/lines. To this end, we represent each template as a field model and matching a template with an input image is performed through estimation of a maximum a posteriori (MAP) in the model. For an effective estimation, variational mean field method is adopted. Variational approach is often used when the exact solution is not feasible or practical to obtain. Its robustness has also been verified in various computer vision tasks, e.g. object detection and tracking [24, 11, 18, 20]. In our proposed template matching method, the variational mean field method is used to infer the locations of edge pixels to be considered as the matching points of the template.

We note that the proposed method differs from the snake model in [14] and active shape model in [5]. In particular, the deformation of the object shape (represented by edge pixels) is controlled by the templates. Furthermore, the objective function of the model can be optimised locally using the variational mean field method, thus the computational complexity can be significantly reduced. Our model also differs from that proposed in [25]. Specifically, we use the DT to efficiently compute the likelihood. In addition, the matching is performed through the MAP estimation while it was done in a hierarchical manner in [25].

The proposed method was applied for two different versions of Chamfer template matching: Oriented Chamfer Matching (OCM) [22, 19] and Directional Chamfer Matching (DCM) [16]. We extensively evaluated the proposed method in the task of object detection on the ETHZShapeClass and INRIA Horse dataset. Experimental

<sup>1</sup>a template can be represented as a set of points [9] or lines [16]

results have shown the advantages of the proposed method in comparison to the DCM, OCM and other object detection techniques.

In the following, the Chamfer template matching with its variants, which provide a background for the proposed method, are briefly presented in section 2. A new formulation of template matching is presented in section 3. Variational mean field method is then described in section 4. Section 5 shows experimental results. The paper is concluded in section 6.

## 2. Background

Let  $I$  denote an input image and  $E(I)$  be its edge map generated using some edge detector, e.g. Canny's detector [3]. On  $E(I)$  a Distance Transform (DT) calculating the distance of every pixel  $t$  to its closest edge pixel in  $E(I)$  is defined as

$$D(t) = \min_{e \in E(I)} \|t - e\|_2 \quad (1)$$

In [19], the authors proposed to use edge magnitude to weight the DT so that strong edge points have more influence than weak edge points which often represent background noise. In particular,  $D(t)$  is modified as

$$D(t) = \min_{e \in E(I)} \left\{ \|t - e\|_2 + \frac{\eta}{\sqrt{\left[\frac{\partial I}{\partial x}(e)\right]^2 + \left[\frac{\partial I}{\partial y}(e)\right]^2}} \right\} \quad (2)$$

where  $\frac{\partial I}{\partial x}(e)$  and  $\frac{\partial I}{\partial y}(e)$  are the horizontal and vertical gradients of the image  $I$  at position  $e$ ,  $\eta$  is a positive constant controlling the contribution of the edge magnitude at  $e$ . Using the method in [6],  $D(t)$  can be computed in  $O(2N)$  where  $N$  is the image size (in pixel).

Let  $T = \{t_1, t_2, \dots, t_{|T|}\}$  be a template in which  $t_i = (x_i, y_i, o_i)$ ,  $i \in \{1, \dots, |T|\}$  includes the location  $(x_i, y_i)$  and orientation  $o_i$  of the template point  $t_i$ ;  $|T|$  is the cardinality of  $T$ . Let  $e(t_i)$  be the closest edge point of  $t_i$  in  $E(I)$ , e.g.  $\|e(t_i) - t_i\|_2 = D(t_i)$  if (1) is used. Note that,  $e(t)$  and  $D(t)$  can be computed simultaneously for all pixel locations. The oriented Chamfer distance at  $t_i$  is defined as

$$d(t_i) = \left\{ [D(t_i)]^n + \lambda [g(o_i, o_{e(t_i)})]^n \right\}^{\frac{1}{n}} \quad (3)$$

where  $o_{e(t_i)}$  is the orientation of  $e(t_i)$ ,  $g(o_i, o_{e(t_i)})$  is some measure of the difference between  $o_i$  and  $o_{e(t_i)}$ ,  $\lambda$  is a weight factor. In [22],  $g(o_i, o_{e(t_i)}) = \min\{|o_i - o_{e(t_i)}|, |o_i - o_{e(t_i)}| - \pi|\}$  and in [19],  $g(o_i, o_{e(t_i)}) = \sin|o_i - o_{e(t_i)}|$ . In addition,  $n$  was set to 1 in [22] and 2 in [19].

In [16], the orientation of edge pixels was integrated in computing the DT. In particular,

$$d(t_i) = \min_{\phi \in \Phi} \{D_\phi(t_i) + \lambda g(\hat{o}_i, \phi)\} \quad (4)$$

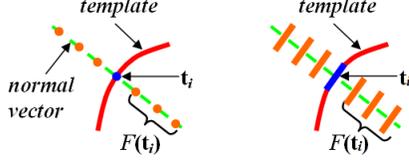


Figure 2. An example of extended points (left) and lines (right).

where  $\Phi$  is the quantised range of orientations,  $D_\phi$  is the DT created for edge points whose orientation is  $\phi$ , and  $\hat{o}_i \in \Phi$  is the nearest quantised orientation of  $o_i$ .

Note that both (1) and (2) can be used to compute  $D_\phi$ . As presented in [16],  $d$  in (4) can be computed once for all pixel locations using dynamic programming with a computational cost of at most  $O(2|\Phi|N)$  where  $|\Phi|$  and  $N$  is the number of quantised orientations and image size (in pixel).

The oriented Chamfer distance between the input image  $I$  and the template  $T$  is finally defined as

$$C(I, T) = \sum_{i=1}^{|T|} d(t_i) \quad (5)$$

where  $d(t_i)$  can be computed using either (3) or (4).

To further save the computational cost in calculating  $C(I, T)$ , Liu et al. [16] proposed matching line segments instead of pixels. In addition, integral images were employed to pre-calculate the distance of pixels on line segments given the segments' orientations. Template matching using (3) and (4) is referred to as oriented Chamfer matching (OCM) [22, 19] and directional Chamfer matching (DCM) [16] respectively.

### 3. Problem Formulation

This section devises a new form of template matching based on Markov random field (MRF) model and conventional Chamfer template matching. Let  $I$ ,  $E(I)$ , and  $T$  be an input image, its edge map, and a template respectively. To cope with the local deformations of the object shape, we allow every template point  $t_i \in T$  to have more than one matching edge point on  $E(I)$ . Specifically,  $T$  is extended by adding sets of points  $F(t_i)$  along the normal vector of template points  $t_i$ . Note that every point in  $F(t_i)$  has the same orientation of  $t_i$ . When  $t_i$  are line segments as in [16],  $F(t_i)$  will include lines parallel to  $t_i$ . Figure 2 shows an example of a template and its extension.

For the sake of simplicity, we assume that  $t_i$  represents template points hereafter as line segments can be applied similarly. The two-layer field model of a template  $T$  is constructed as follows. For each  $t_i \in T$ , let  $h_i$  and  $v_i$  be the hidden and observation node respectively;  $h_i$  takes values in  $F(t_i) \cup \{t_i\}$ ,  $v_i$  is the closest edge point of  $h_i$  on the edge map  $E(I)$ . Note that for every  $h_i$ ,  $v_i$  can be determined with

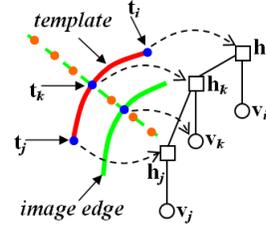


Figure 3. An example of the field model.

a constant complexity using the DT computation method proposed in [6]. On the hidden layer, each hidden node  $h_i$  is linked to its observation node  $v_i$  by an undirected edge. The hidden node  $h_i$  is also directly connected to other hidden nodes  $h_j$  whose distance to  $h_i$  is less than a radius  $r$ . Figure 3 shows the field model of a template.

Given  $T$ ,  $H(T) = \{h_i\}, i \in \{1, \dots, |T|\}$  can be determined as above, the matching cost  $M(I, T)$  between the image  $I$  and template  $T$  can be considered as the similarity between the set of template points  $H(T)$  and the subset of edge points of  $E(I)$  that best fits  $H(T)$ . Similarly to the conventional Chamfer template matching, the computation of  $M(I, T)$  can be relaxed to calculating the fitness of  $H(T)$  to the edge map  $E(I)$ . Since  $H(T)$  covers local but regular deformations of  $T$ , the problem becomes to find a configuration of  $H(T)$  that best fits  $E(I)$ . In other words, this corresponds to compute a maximum of a posteriori (MAP)  $p(H(T)|V = \{v_i\})$  over all possible configurations of  $H(T)$ , i.e.,

$$M(I, T) = \max_{H(T)} p(H(T)|V) = \max_{H(T)} \frac{p(V|H(T))p(H(T))}{p(V)} \quad (6)$$

Let  $p(v_i|h_i)$  be the likelihood of having an edge point  $v_i$  given a template point  $h_i$ . Assume that  $p(v_i|h_i)$  can be computed using some distance, e.g.,

$$p(v_i|h_i) \propto \exp[-\alpha d(h_i)] \quad (7)$$

where  $d(h_i)$  is computed as in (3) or (4),  $\alpha$  is a positive parameter.

We further assume that: 1)  $v_i$  is independent of each other, i.e.  $p(V|H(T)) = p(v_1, v_2, \dots, v_{|H(T)|}|H(T)) = \prod_{i=1}^{|H(T)|} p(v_i|H(T))$ , and 2)  $v_i$  is determined based on only  $h_i$  using the DT, i.e.  $p(v_i|H(T)) = p(v_i|h_i), \forall i \in \{1, \dots, |H(T)|\}$ , and 3)  $p(V)$  is uniform. The MAP problem in (6) can be rewritten as,

$$M(I, T) \propto \max_{H(T)} \prod_{i=1}^{|H(T)|} p(v_i|h_i)p(H(T)) \quad (8)$$

Note that the OCM in [22, 19] and DCM in [16] can be considered as special cases of our proposed template match-

ing where extended template points are not used and every template point is matched independently. Indeed, when  $F(t_i) = \emptyset, i \in \{1, \dots, |T|\}$ ,  $H(T)$  becomes  $T$ , i.e.  $h_i = t_i$ , and  $V$  is the set of closest edge points of  $T$ . In addition, assuming that  $p(T)$  is uniform and using (7) and (5), (8) can be simplified as,

$$\begin{aligned} M(I, T) &\propto \prod_{i=1}^{|T|} p(v_i | t_i) \\ &= \exp \left[ -\alpha \sum_{i=1}^{|T|} d(t_i) \right] = \exp[-\alpha C(I, T)] \quad (9) \end{aligned}$$

where  $C(I, T)$  is defined in (5).

As can be seen in (8), the term  $p(H(T))$  is the prior of possible locations of template points. It represents the constraint on the deformations of the object shape. To estimate  $M(I, T)$  using (8), an exhausted search over all possible configurations of  $H(T)$  would require  $O(\prod_{i=1}^{|T|} |F(t_i) \cup \{t_i\}|) = O(\prod_{i=1}^{|T|} (|F(t_i)| + 1))$  operations. Assume that  $|F(t_i)| + 1 = \mathbf{f}$  for every  $t_i \in T$ , the computational complexity of the estimation of  $M(I, T)$  using (6) would be  $O(\mathbf{f}^{|T|})$ . In addition, since the MRF model is not a tree structure, exact inference algorithms such as dynamic programming (or Viterbi) [23], belief propagation cannot be applied. In the following section, an alternative solution to effectively estimate  $M(I, T)$  based on the variational mean field approach will be proposed.

#### 4. Variational Mean Field Approach

For simplicity but without ambiguity,  $H$  will be used instead of  $H(T)$  hereafter. The core idea of the variational approach in estimation of a MAP  $p(H|V)$  is to use an analytical but simple variational distribution  $Q(H)$  to approximate  $p(H|V)$  and at the same time to approximate  $\log p(V)$  through optimising an objective function  $J(Q)$  as follows,

$$\begin{aligned} J(Q) &= \log p(V) - KL(Q(H)||p(H|V)) \\ &= - \int_H Q(H) \log Q(H) dH + \int_H Q(H) \log p(H, V) dH \quad J(Q) = \text{const.} + \mathcal{H}(Q_i) + \int_{h_i} Q_i(h_i) E_Q \{ \log p(H, V) | h_i \} \\ &= \mathcal{H}(Q) + E_Q \{ \log p(H, V) \} \quad (10) \end{aligned}$$

where  $\mathcal{H}(Q)$  is the entropy of the variational distribution  $Q$ ,  $E_Q \{ \cdot \}$  represents the expectation with regard to  $Q$ ,  $KL$  is the Kullback-Leibler divergence [15] defined as,

$$KL(Q(H)||p(H|V)) = \int_H Q(H) \log \frac{Q(H)}{p(H|V)} dH \quad (11)$$

As shown in (10),  $J(Q)$  is the lower bound of  $\log p(V)$  (as the KL-divergence is nonnegative). Thus, maximising  $J(Q)$  with respect to  $Q$  corresponds to calculating the optimal approximation of both  $\log p(V)$  and  $p(H|V)$ . In this

paper, the simplest variational distribution which can be fully factorised is adopted as,

$$Q(H) = \prod_{i=1}^{|T|} Q_i(h_i) \quad (12)$$

where  $Q_i(h_i)$  is the distribution of  $h_i$ .

The entropy  $\mathcal{H}(Q)$  then becomes,

$$\mathcal{H}(Q) = \sum_{i=1}^{|T|} \mathcal{H}(Q_i) \quad (13)$$

where  $\mathcal{H}(Q_i)$  is the entropy of  $Q_i$ .

As shown in [12], the optimum of  $J(Q)$  can be obtained by a set of interrelated Gibbs distributions:

$$Q_i(h_i) = \frac{1}{Z_i} e^{E_Q \{ \log p(H, V) | h_i \}} \quad (14)$$

where  $E_Q \{ \cdot | h_i \}$  is the conditional expectation with respect to the variational distribution  $Q$  given  $h_i$ ,  $Z_i$  is the normalisation factor computed as,

$$Z_i = \int_{h_i} e^{E_Q \{ \log p(H, V) | h_i \}} \quad (15)$$

In addition, the maximisation of  $J(Q)$  can be performed individually for each  $Q_i$ , i.e.  $Q_j, j \neq i$  remain unchanged when  $Q_i$  is updated using (14). In other words,

Equations (14) and (15) will be called iteratively until the optimum value of  $J(Q)$  is obtained using (16). To compute  $Q_i(h_i)$ , it is required to estimate  $E_Q \{ \log p(H, V) | h_i \}$ . As in a standard MRF, we assume that the estimation of  $E_Q \{ \log p(H, V) | h_i \}$  depends only on neighbouring sites of  $h_i$ . In particular, let  $\mathcal{N}(h_i)$  denote the set of hidden neighbours of  $h_i$ . As presented in [13], the update of  $E_Q \{ \log p(H, V) | h_i \}$  (and also  $J(Q)$ ) can be done locally on *cliques* (or *edges* in the field model) containing  $h_i$  as,

$$\begin{aligned}
& E_Q \{ \log p(H, V) | h_i \} \\
& \leftarrow E_Q \left\{ \log \left[ p(v_i, h_i) \prod_{h_j \in \mathcal{N}(h_i)} p(h_i, h_j) \right] \right\} \\
& = \log p(v_i | h_i) + \log p(h_i) \\
& + E_Q \left\{ \sum_{h_j \in \mathcal{N}(h_i)} \log p(h_j) + \log p(h_i, h_j) \right\} \\
& = \log p(v_i | h_i) + \log p(h_i) \\
& + \sum_{h_j \in \mathcal{N}(h_i)} \int_{h_j} Q_j(h_j) \log p(h_j) \\
& + \sum_{h_j \in \mathcal{N}(h_i)} \int_{h_j} Q_j(h_j) \log p(h_i, h_j) \quad (17)
\end{aligned}$$

where the likelihood  $p(v_i | h_i)$  is computed similarly to (7).

In (17),  $p(h_i)$  and  $p(h_i, h_j)$  can be considered as the potential functions of a MRF. Assume that every template point has the same importance,  $p(h_i)$  can be set to a constant;  $p(h_i, h_j)$  is computed as,

$$p(h_i, h_j) \propto \exp \left[ -\beta |\Theta(\overrightarrow{h_i h_j}, \overrightarrow{t_i t_j})| \right] \quad (18)$$

where  $\beta$  is a user-defined value,  $\Theta(\overrightarrow{h_i h_j}, \overrightarrow{t_i t_j})$  is some measure of the angle between two vectors  $\overrightarrow{h_i h_j}$  and  $\overrightarrow{t_i t_j}$  (e.g.  $\Theta(\overrightarrow{h_i h_j}, \overrightarrow{t_i t_j}) = 1 - |\cos(\overrightarrow{h_i h_j}, \overrightarrow{t_i t_j})|$  in our implementation);  $t_i$  and  $t_j$  are template points (i.e.  $h_i \in F(t_i) \cup \{t_i\}$  and  $h_j \in F(t_j) \cup \{t_j\}$ ).

If  $h_i$  and  $h_j$  are line segments as used in [16],  $\overrightarrow{h_i h_j}$  can be computed as the vector connecting the middle points of  $h_i$  and  $h_j$ . As can be seen, the term  $p(h_i, h_j)$  encodes the local deformations of the template and is compensated by the likelihood  $p(v_i | h_i)$  computed individually on every template point/line.

To update  $E_Q \{ \log p(H, V) | h_i \}$  using (17), we assume that  $Q_i(h_i)$  is initialised uniformly, i.e.  $Q_i(h_i) = \frac{1}{|F(t_i) \cup \{t_i\}|} = \frac{1}{|F(t_i)| + 1}$ . Finally, after  $J(Q)$  is maximised, the returned variational distribution  $Q(H)$  can be considered as an approximate of  $p(H|V)$ . In addition, since  $Q(H)$  is fully factorised,  $M(I, T)$  in (8) can be estimated as,

$$M(I, T) \approx \prod_{i=1}^{|T|} Q_i(h_i^*) \quad (19)$$

where  $h_i^* = \arg \max_{h_i} Q_i(h_i)$ .

Assume that each node  $h_i$  has  $\mathbf{n}$  edges connecting it to other hidden nodes and to  $v_i$ ,  $E_Q \{ \log p(H, V) | h_i \}$  in (17) can be updated in  $O(\mathbf{nf}^2)$  where  $\mathbf{f} = |F(t_i)| + 1$ . Thus, the computational complexity of matching all points on a

template  $T$  would be  $O(\mathbf{nf}^2 |T|)$ . This is the advantage of the variational approach compared with the brute-force estimation of  $\max_H p(H|V)$  that requires the complexity of  $O(\mathbf{f}^{|T|})$  for searching all possible configurations of  $H$ .

## 5. Experimental Results

### 5.1. Experimental Setup

The proposed method was applied for object detection. We experimented the proposed method with the use of two common Chamfer template matching techniques: OCM [22, 19] and DCM [16] in calculating the distance values. In particular, we computed  $d(h_i)$  in (7) using (3) and (4). Recall that when DCM [16] is used, each  $h_i$  is a line segment and  $d(h_i)$  is the sum of the distance values of all points on  $h_i$ . To speed-up the computation, integral images corresponding to different directions were used. For the DCM method, we used the default settings provided by the authors in their paper, e.g. the number of scales was 8, the ratio between two consecutive scales was 1.2, the same non-maximal suppression (to merge overlapping detection results) was used. For the OCM method [22, 19],  $\lambda$  was set to 1.0. For the variational mean field model,  $\alpha$  in (7) and  $\beta$  in (18) were set to 10.0 and 5.0 respectively.

The proposed method was tested on two datasets: ETHZShapeClass [8] and INRIA Horse [7]. Both the datasets are challenging to include objects in various sizes, appearance, and high articulation. Edge maps are available in both the test sets. The ETHZShapeClass dataset includes 255 images containing five different classes: apple logos, bottles, giraffes, mugs, and swans. On this set, templates (one per class) are also provided. Once an object class is evaluated, images of other classes are considered as negative images. The INRIA Horse dataset consists of 170 images containing instances of horse and other 170 background images. This set does not include any templates, thus we manually created a template for the experiments.

### 5.2. Performance Evaluation

We first evaluated the proposed method in two cases: with the use of OCM [22, 19] and DCM [16] in calculating the distance values. For each case, we investigated the detection accuracy when the number of extended points/lines for each template point/line  $t_i$  was varied in 1, 2, and 3 for each direction of the normal vector. This means that the total number of extended points/lines per template point/line will be 2, 4, and 6; and the number of possible values that each  $h_i$  can take will be 3, 5, and 7 points/lines respectively. In our experiments, extended points/lines were distributed uniformly along the normal vectors and the interval between adjacent extended points/lines was set to 3 (pixels). Note that this value could be set adaptively to the template size. Figure 4 shows the detection performance of the pro-

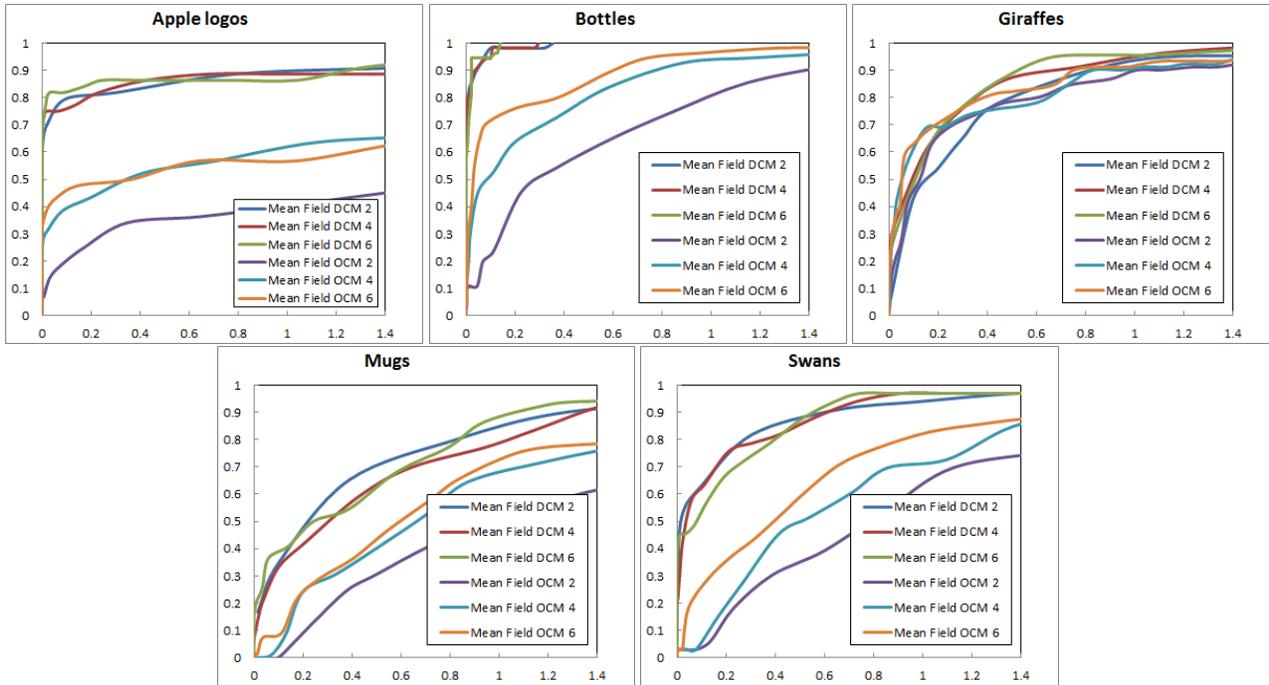


Figure 4. The ROC curves of variants of the proposed method, e.g. Mean Field DCM 2 indicates the use of DCM to calculate the distances in (7) and the number of extended points/lines for each template point/line is 2. The horizontal axis corresponds to False Positive Per Image rate and the vertical axis represents the Detection rate. This figure is best viewed in colour.

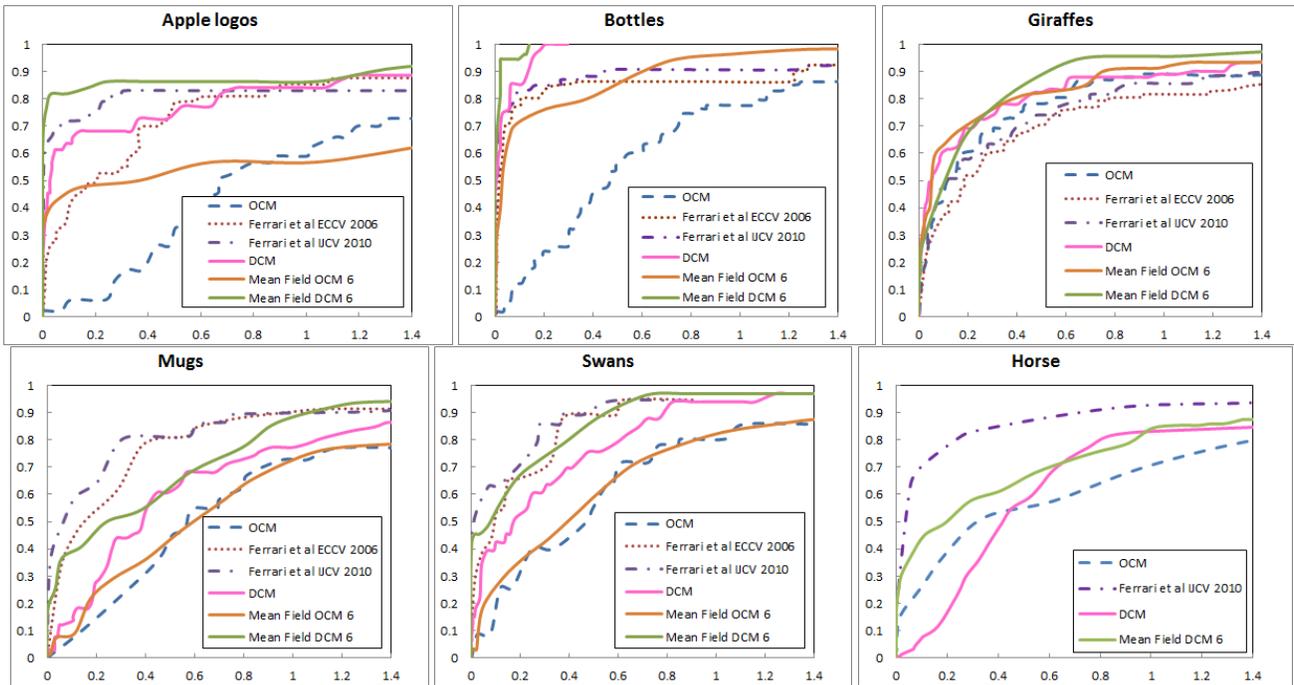


Figure 5. Comparison of the proposed method with other existing methods. This figure is best viewed in colour.

posed method when OCM and DCM was used with different number of extended points/lines in the template. As can

be seen in this figure, the different settings gain different performances. We have also noticed that, except the “Gi-



Figure 6. Some detection results of the apple logos (1st row), bottles (2nd row), giraffes (3rd row), mugs (4th row), swans (5th row) and horses (6th row). Templates are the most right images.

raffes” class, in all other cases, the DCM significantly outperformed the OCM. Some detection results are presented in Figure 6.

We have also investigated the computational complexity of the proposed template matching method. For the best performance (with DCM), we have found that, on the average, one image could be processed in 0.76 seconds approximately. These experiments were conducted on the ETHZShapeClass dataset and on an Intel(R) Core(TM) i7 2.10GHz CPU computer with 8.00 GB memory.

### 5.3. Comparison

In addition to performance evaluation, we compared the proposed method with existing methods including [22] (marked as “OCM”), [16] (marked as “DCM”), [8] (marked

as “Ferrari et al ECCV 2006”), and [7] (marked as “Ferrari et al IJCV 2010”). Figure 5 shows the comparison results on the ETHZShapeClass and INRIAHorse dataset. As can be seen in this figure, in general, the use of variational mean field improves the detection accuracy when it is applied to the OCM and DCM. On the ETHZShapeClass, the proposed method could obtain comparable performance in comparison to the state-of-the-arts. On the INRIAHorse dataset, the method in [7] outperformed our method. However, it is notice that only one template was used in our method. It would be expected that better performance could be gained if more templates are used. Moreover, our method does not require off-line training and thus it is suitable for applications where the templates are provided by the user on the fly.

For the computational complexity, as reported in [16], the DCM method could achieve roughly 0.39 seconds per image by truncating more than 90% of detection hypotheses. In our experiments, we accepted more detection hypotheses to avoid miss detections. However, experimental results have shown that the proposed method still kept low false alarm rate in comparison to the DCM method.

Although the methods in [8, 7] did not report the processing time of their detection system, they potentially have high computational complexity. First, extracting pairs of adjacent segments from edge maps requires some level of computations. Second, those methods make use the Hough transform to locate the object; while Hough transform is known for its highly computational complexity. Third, the Thin-Plate Spline Robust Point Matching algorithm [4] is used to refine the detection results. Again, this algorithm is expensively computational as acknowledged by the authors.

## 6. Conclusion

This paper proposes a novel mean field-based template matching method. In the proposed method, the template is represented as a field model in which hidden variables correspond to possible locations of the template points and observation nodes are their closest edge points. The problem of template matching is then formulated as estimation of a maximum a posteriori (MAP) of hidden variables given the observation data. Mean field variational method is adopted in the paper to effectively approximate the MAP. The proposed method was applied to two common Chamfer template matching techniques for the task of object detection. Experimental results on two challenging datasets have showed that the proposed method significantly improved the detection accuracy in comparison to the two Chamfer template matching techniques and achieved comparable performance to state-of-the-art on the test sets.

## Acknowledgements

The author would like to thank the reviewers for constructive comments and A/Prof. Wanqing Li for helping revising the manuscript.

## References

- [1] X. Bai, Q. Li, L. J. Latecki, W. Liu, and Z. Tu. Shape band: A deformable object detection approach. In *CVPR*, pages 1335–1342, 2009.
- [2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24(4):509–522, 2002.
- [3] J. F. Canny. A computational approach to edge detection. *PAMI*, 8(6):679–698, 1986.
- [4] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *CVIU*, 89(2-3):114–141, 2003.
- [5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *CVIU*, 61(1):38–59, 1995.
- [6] P. F. Felzenszwalb and D. P. Huttenlocher. Distance transforms of sampled functions. Technical report, Cornell Computing and Information Science, 2004.
- [7] V. Ferrari, F. Jurie, and C. Schmid. From images to shape models for object detection. *IJCV*, 87(3):284–303, 2010.
- [8] V. Ferrari, T. Tuytelaars, and L. V. Gool. Object detection by contour segment networks. In *ECCV*, pages 14–28, 2006.
- [9] D. M. Gavrilu. Multi-feature hierarchical template matching using distance transforms. In *ICPR*, volume 1, pages 439–444, 1998.
- [10] D. M. Gavrilu. A Bayesian, exemplar-based approach to hierarchical shape matching. *PAMI*, 29(8):1–14, 2007.
- [11] G. Hua and Y. Wu. A decentralized probabilistic approach to articulated body tracking. *CVIU*, 108:272–283, 2007.
- [12] T. S. Jaakkola. Tutorial on variational approximation methods. Technical report, MIT Artificial Intelligence Laboratory, 2000.
- [13] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. Saul. An introduction to variational methods for graphical models. *Machine Learning*, pages 183–233, 1999.
- [14] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, pages 321–331, 1988.
- [15] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:76–86, 1951.
- [16] M. Y. Liu, O. Tuzel, A. Veeraraghavan, and R. Chellappa. Fast directional chamfer matching. In *CVPR*, pages 1696–1703, 2010.
- [17] T. Ma, X. Yang, and L. J. Latecki. Boosting chamfer matching by learning chamfer distance normalization. In *ECCV*, volume 5, pages 450–463, 2010.
- [18] C. Medrano, J. E. Herrero, J. Martínez, and C. Orrite. Mean field approach for tracking similar objects. *CVIU*, 113:907–920, 2009.
- [19] D. T. Nguyen, W. Li, and P. Ogunbona. An improved template matching method for object detection. In *ACCV*, volume 3, pages 193–202, 2009.
- [20] D. T. Nguyen, W. Li, and P. Ogunbona. Inter-occlusion reasoning for human detection based on variational mean field. *Neurocomputing*, 110:51–61, 2013.
- [21] C. F. Olson and D. P. Huttenlocher. Automatic target recognition by matching oriented edge pixels. *IEEE Trans. Image Processing*, 6(1):103–113, 1997.
- [22] J. Shotton, A. Blake, and R. Cipolla. Multiscale categorical object recognition using contour fragments. *PAMI*, 30(7):1270–1281, 2008.
- [23] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla. Shape context and chamfer matching in cluttered scenes. In *CVPR*, volume 1, pages 127–133, 2003.
- [24] Y. Wu and T. Yu. A field model for human detection and tracking. *PAMI*, 28(5):753–765, 2006.
- [25] L. Zhu and A. Yuille. A hierarchical compositional system for rapid object detection. In *NIPS*, 2005.