

Deformable Object Matching via Deformation Decomposition based 2D Label MRF

Kangwei Liu, Junge Zhang, Kaiqi Huang and Tieniu Tan
Center for Research on Intelligent Perception and Computing (CRIPC)
National Laboratory of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences (CASIA)
{kwliu, jgzhang, kqhuang, tnt}@nlpr.ia.ac.cn

Abstract

Deformable object matching, which is also called elastic matching or deformation matching, is an important and challenging problem in computer vision. Although numerous deformation models have been proposed in different matching tasks, not many of them investigate the intrinsic physics underlying deformation. Due to the lack of physical analysis, these models cannot describe the structure changes of deformable objects very well. Motivated by this, we analyze the deformation physically and propose a novel deformation decomposition model to represent various deformations. Based on the physical model, we formulate the matching problem as a two-dimensional label Markov Random Field. The MRF energy function is derived from the deformation decomposition model. Furthermore, we propose a two-stage method to optimize the MRF energy function. To provide a quantitative benchmark, we build a deformation matching database with an evaluation criterion. Experimental results show that our method outperforms previous approaches especially on complex deformations.

1. Introduction

Deformable object matching is a fundamental and challenging problem in computer vision. It has a wide range of applications such as medical image analysis [12], handwriting recognition [16] and object classification [9]. Given two images as illustrated in Fig. 1, usually one is called *reference image* and the other one is called *target image*. The objective of the problem is to obtain the correspondence of features or structures between the two images.

The non-rigid deformations of objects are complex and cannot be handled by calculating the transformation matrix, or simply applying geometric constraints. To solve the problem, the basic idea is to distort the reference image to match the target image. Thus, there are two important aspects in deformation models: 1) how to measure the appear-

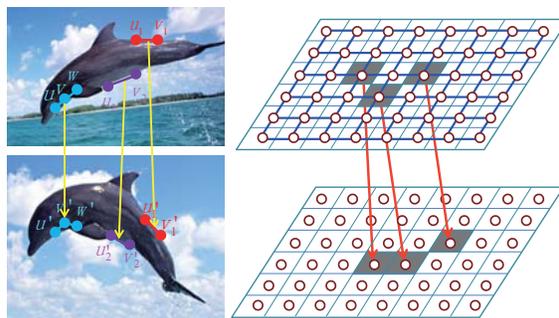


Figure 1. The proposed deformation matching model. The right image shows the 2D label MRF framework. The left image shows the physical analysis of deformation. The match $(u, v, w) \Rightarrow (u', v', w')$, $(u_2, v_2) \Rightarrow (u'_2, v'_2)$ correspond to bending and compression deformation, respectively. There is no internal deformation when (u_1, v_1) is distorted to (u'_1, v'_1) .

ance similarity; 2) how to describe object deformation and preserve the inherent structure of the distorted object. Usually, the former one depends on the characteristic of images in different applications, and the latter is the study emphasis in deformation models.

Numerous deformation models have been proposed for decades. Most models [8, 9, 16, 28] formulate the problem as an energy optimization problem, which simultaneously minimizes the appearance similarity term while satisfying some manually defined smooth constraints to keep the continuity. However, such constraints cannot handle large and complex deformations very well and hence the models will fail to keep the structure of the distorted object. On the other hand, objects (*e.g.*, human pose and face) with different physical properties react in very different ways of deformation. The models cannot describe different structure changes of objects, since they do not take the physical properties into account. The pictorial structures models [10, 11], which can be treated as physics based models, apply springs to represent the object deformation. However, they only get the match of several parts and cannot address accurate cor-

responsiveness of local features. Meanwhile, the springs in the models cannot describe the bending deformation of objects. Jain *et al.* [18] propose a rubber plate model to represent deformations by distorting the template. However, its ability to describe deformations is limited, since the rubber plate cannot be deformed largely (*e.g.*, the deformations shown in Fig. 1). Besides, only the correspondence of object’s edges or contours can be obtained in [18]. Therefore, we believe that the physical analysis of deformation mechanism and an effective framework are important for deformable object matching. Unfortunately, in spite of the existence of physical models in previous literature, most of them exist in graphics [13] and medical image analysis [24], and there are few physical models for deformable object matching.

Motivated by this, we analyze the object deformation mechanism and propose a novel deformation decomposition model (DDM) to describe various deformations. Although the non-rigid deformations of objects are complex, we consider that various deformations can be decomposed into three basic types of deformation: *tension*, *compression* and *bending*, as illustrated in Fig. 2. In this way, complex deformation can be easily addressed by deformation decomposition. The proposed DDM is composed of two lattices, which describe axial deformation (tension and compression) and bending deformation, respectively. Our model is similar to the work of Chenchiah *et al.* [5], in which the deformation energy is also decomposed into two kinds of energies. However, there is no clear physical meanings of their two kinds of energies.

Based on the DDM, we formulate the deformable object matching problem as an energy optimization task of *two-dimensional label Markov Random Field* as illustrated in Fig. 1. Furthermore, we propose a generalized framework of the 2D label MRF and extend it to *n-dimensional label MRF*. Finally, we propose a two-stage method to optimize the MRF energy which is derived from the DDM.

To provide a quantitative benchmark, we construct a database focusing on the problem of deformable object matching with an effective evaluation criterion. We test our approach on both the constructed database and deformable image pairs from [22]. Furthermore, we apply the DDM on handwriting recognition, which is one of the most important applications of deformation models. Experimental results show that our method outperforms previous models.

There are three main contributions in this paper:

1. We analyze the deformation mechanism and propose a deformation decomposition model. Experiments show that the proposed model can describe various and complex deformations effectively.
2. We propose a 2D label MRF framework for deformable object matching based on the DDM. The framework addresses the correspondence of objects involving complex deformations and is flexible for dif-

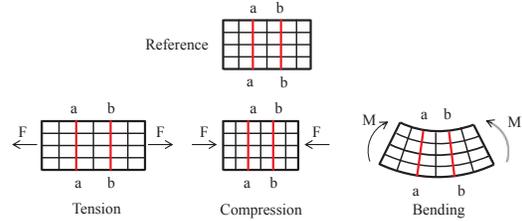


Figure 2. The three basic types of deformation in elastic object: *tension*, *compression* and *bending*.

ferent applications. Moreover, the framework can be easily extended to handle high dimensional cases involved in many vision tasks with a more generalized *n-dimensional label MRF*.

3. We construct a database for deformable object matching and propose an evaluation criterion. The constructed database is direct and effective to compare the matching performance of different deformation models.

The rest of this paper is organized as follows. In Sec. 2, we analyze the deformation mechanism and introduce the proposed DDM. Based on the DDM, we propose the 2D label MRF framework for deformable object matching in Sec. 3. In Sec. 4, we apply a two-stage method to optimize the MRF energy function. Experiments and applications are shown in Sec. 5. We make conclusions in Sec. 6.

2. The analysis and modeling of deformation

In mechanics, deformation is the transformation of an object from a reference configuration to the current configuration [29]. Usually, it occurs when some parts or the whole body of objects are stretched, compressed or warped. There will be internal forces between the connected parts in the deformable object. According to mechanics of materials [2, 3], there are three principal internal forces: tension stress (stress is the force per unit area inside the object), compression stress and bending moment, and the corresponding basic types of deformation are tension, compression and bending (see Fig. 2). Since deformations can be decomposed into the three basic deformations, complex deformations can be easily addressed by deformation decomposition.

Motivated by this, we propose the deformation decomposition model (DDM). The DDM is composed of a tension-compression spring lattice (T-lattice) and a bending spring lattice (B-lattice) as shown in Fig. 3. The two lattices describe the axial and bending deformation respectively.

Definition 2.1 The tension-compression spring lattice (T-lattice) is defined as $(\mathcal{Z}^{M \times N}, \mathcal{L}_t)$ where

1. \mathcal{Z} is the set of non-negative integers, $\mathcal{Z}^{M \times N}$ is the two dimensional integer sequence $\{1, \dots, M\} \times \{1, \dots, N\}$ and the nodes in the lattice are points in $\mathcal{Z}^{M \times N}$.

2. \mathcal{L}_t is the set of the initial length l_0 of the tension-compression spring.

In the T-lattice, nodes $u, v \in \mathcal{Z}^{M \times N}$ are connected by a spring, if $l_{uv} \in \mathcal{L}_t$ where l_{uv} denotes the distance between u and v . For brevity, we use (u, v) to denote the spring connecting u and v . In the T-lattice, only the springs whose initial length $l_{uv} = 1$ are applied, that is $\mathcal{L}_t = \{1\}$.

When deformation occurs, there will be deformation energy stored in the object due to the work done by stresses. The tension-compression potential (T-potential) can be calculated according to the work done by springs in the T-lattice. With Hooke's law $F_t = k_t \Delta l$, the potential stored in the spring (u, v) is

$$T_{uv} = \int_0^{\Delta l} F_t d\delta = \frac{1}{2} k_t (l'_{uv} - l_{uv})^2 \quad (1)$$

where k_t denotes the tension and compression modulus representing the object's physical property in axial direction, l_{uv} is the initial length of the spring (u, v) , l'_{uv} is the current length after deformation and Δl represents the length changes of the spring. If $\Delta l > 0$, F_t is the tension stress and otherwise F_t is the compression stress.

Thus, the T-potential on the T-lattice can be expressed as

$$E_t = \sum_{l_{uv} \in \mathcal{L}_t} T_{uv} = \sum_{l_{uv} \in \mathcal{L}_t} \frac{1}{2} k_t (l'_{uv} - l_{uv})^2 \quad (2)$$

In spite of its capability to describe axial deformation effectively, the T-lattice cannot represent bending deformation. Therefore, it is necessary to apply a lattice to describe the bending deformation.

Definition 2.2 The bending spring lattice (B-lattice) is defined as $(\mathcal{Z}^{M \times N}, \vartheta_b)$, where ϑ_b is the set of initial angle of bending springs.

In the B-lattice, the bending springs can be bent and will store energy due to the bending deformation. Nodes $u, v, w \in \mathcal{Z}^{M \times N}$ are connected by a bending spring if $\theta_{uvw} \in \vartheta_b$, where θ_{uvw} denotes the angle between (u, v) and (v, w) . In the B-lattice, only the bending springs whose initial angle $\theta_{uvw} = \pi$ are applied, that is $\vartheta_b = \{\pi\}$.

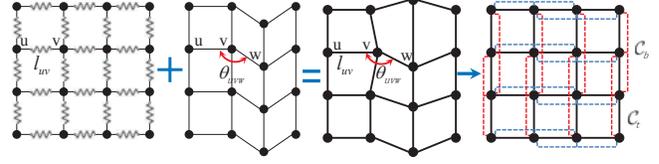
The bending potential (B-potential) can be calculated by the work done by bending springs. The work done by the bending spring (u, v, w) is

$$B_{uvw} = \int_0^{\Delta \theta} M_b d\theta = \frac{1}{2} k_b (\theta'_{uvw} - \theta_{uvw})^2 \quad (3)$$

where k_b is the bending modulus which represents the bending properties, θ_{uvw} is the initial angle of (u, v, w) , θ'_{uvw} is the current angle after deformation and M_b is the bending moment of the bending spring.

Thus, the bending potential on the whole lattice can be expressed as

$$E_b = \sum_{\theta_{uvw} \in \vartheta_b} B_{uvw} = \sum_{\theta_{uvw} \in \vartheta_b} \frac{1}{2} k_b (\theta'_{uvw} - \theta_{uvw})^2 \quad (4)$$



(a) T-lattice $(\mathcal{Z}^{M \times N}, \mathcal{L}_t)$ (b) B-lattice $(\mathcal{Z}^{M \times N}, \vartheta_b)$ (c) DDM $(\mathcal{Z}^{M \times N}, \mathcal{L}_t, \vartheta_b)$ (d) MRF $\mathcal{G}=(\mathcal{V}, \mathcal{C})$

Figure 3. (c) shows the configuration of DDM after deformation. The DDM is composed of a T-lattice (a) and a B-lattice (b). (d) shows the 2D label MRF model for deformable object matching, which is based on the DDM.

Therefore, the final DDM is a lattice $(\mathcal{Z}^{M \times N}, \mathcal{L}_t, \vartheta_b)$, which is the composition of the T-lattice and the B-lattice (see Fig.3). When deformation occurs, the displacements of nodes in the model can be addressed by minimizing the deformation energy according to the minimum potential energy principle [4]. The deformation energy is composed of tension-compression potential and bending potential. In contrast to traditional methods calculating equations of equilibrium, the energy methods do not need to solve a set of auxiliary quantities (e.g., stresses) of no interest to obtain the required answer [2].

The above model is the basis of our framework. In Sec. 3, we will introduce the framework of 2D label MRF for deformable object matching. The labels in MRF model represent displacements of corresponding nodes in the DDM. Meanwhile, the MRF energy function is derived from the deformation energy. In this way, we transform the minimization of the deformation energy into the optimization of MRF energy function.

3. The framework of deformable object matching

The goal of deformable object matching is to obtain the mapping from the reference image I_r to the target image I_t . In the proposed framework, we divide the image into a set of small rectangle elements. Let \mathcal{T}_r and \mathcal{T}_t denote the sets of elements in I_r and I_t respectively. Each element in \mathcal{T}_r corresponds to a node in the DDM. Therefore, if we get the displacement of each node in the DDM, we can obtain the mapping $\mathcal{M}: \mathcal{T}_r \rightarrow \mathcal{T}_t$ with the relationship

$$(x', y') = (x, y) + (X, Y) \quad (5)$$

where (x, y) is the coordinate of an element in \mathcal{T}_r , (x', y') is the coordinate of the matched element in \mathcal{T}_t and (X, Y) denotes the displacement of element (x, y) .

To optimize the displacements of each node, the deformable object matching problem can be formulated as a 2D label MRF (see Fig. 3), whose labels are two dimensional and represent the displacement of nodes.

3.1. Two-dimensional label MRF

Definition 3.1 Given a MRF graph $\mathcal{G} = (\mathcal{V}, \mathcal{C})$ on the reference image I_r , it consists of

1. A finite set $\mathcal{V} \in \mathcal{Z}^{M \times N}$ of nodes, which represent corresponding elements in \mathcal{T}_r . Each node $v \in \mathcal{V}$ can be denoted by a two dimensional coordinate (x_v, y_v) and there is one-to-one correspondence between the nodes in \mathcal{V} and the nodes in the DDM.
2. A set \mathcal{C} of cliques, each of which is a subset of \mathcal{V} . For example, $(u, v) \in \mathcal{C}$ if node u is connected with v .
3. A label set $\mathcal{L} = \mathcal{X} \times \mathcal{Y}$, which is a two dimensional discrete space. \mathcal{X} and \mathcal{Y} , which are one dimensional, represent the horizontal and vertical displacements of nodes in the DDM respectively:

$$\begin{aligned} \mathcal{X} &= \{-x_m, \dots, x_0, \dots, x_m\} \\ \mathcal{Y} &= \{-y_n, \dots, y_0, \dots, y_n\} \end{aligned} \quad (6)$$

where $\pm x_m, \pm y_n$ are the maximal displacements in x and y directions.

The most important difference between the above model and the multi-label MRF is that the label of each node is a two dimensional variable $(X, Y) \in \mathcal{L}$. The probability distribution on a configuration of the model is the joint probability distributions of variables X and Y . Thus, the model cannot be simply regarded as a multi-label MRF. We call the defined MRF model *two-dimensional label MRF*. Therefore, the mapping $\mathcal{M} : \mathcal{T}_r \rightarrow \mathcal{T}_t$ can be obtained by optimizing the labels of the 2D label MRF model.

In fact, similar 2D label space has been applied in [20, 28, 32]. However, none of them propose a generalized framework of the 2D label MRF and even several papers treat it as a multi-label MRF. In the 2D label MRF, the labels X and Y can represent different statistical or physical meanings depending on different problems. For an instance of simultaneously optimizing stereo and segmentation [20], the model can be regarded as a case of the 2D label MRF, in which X and Y represent the depth and the object class respectively. In contrast, the multi-label MRF cannot solve the two problems simultaneously, since its label can only represent one kind of meaning.

Furthermore, we extend the concept to a more generalized framework of *n-dimensional label MRF* whose label set \mathcal{L} is a n -dimensional discrete space:

$$\mathcal{L} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_n \quad (7)$$

In the n -dimensional label MRF, a configuration $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ assigns a n -dimensional variable to each node, and each dimensional label \mathcal{X}_n has different representations. The multi-label MRF is a special case of n -dimensional label MRF when the label dimension $n = 1$.

3.2. Energy function

The energy of the 2D label MRF contains three terms. The first term is data term and the other two terms are tension-compression potential and bending potential. The data term measures the appearance similarity. Meanwhile, the deformation energy, which is the sum of T-potential and B-potential, preserves the inherent structures of deformable objects in the matching process.

Data term. Each node in the graph \mathcal{G} represents the corresponding image region with an associated feature vector. The appearance feature can be pixel, histogram of oriented gradients [7], or scale-invariant feature transform [23] *etc.* in different tasks. The similarity between the element in \mathcal{T}_r and the corresponding element in \mathcal{T}_t is the distance between the two associated features. The data term can be expressed as

$$E_d(\mathbf{L}) = \sum_{v \in \mathcal{V}} D(L_v) \quad (8)$$

where $\mathbf{L} = (\mathbf{X}, \mathbf{Y}) = \{L_v\}$ is a configuration of the 2D label MRF and assigns to each node v a label $L_v = (X_v, Y_v)$.

Tension and compression potential. Let \mathcal{C}_t denote the set of cliques representing the tension and compression deformation. If $l_{uv} \in \mathcal{L}_t$, $(u, v) \in \mathcal{C}_t$, where $\mathcal{L}_t = \{1\}$ in the DDM. Using Eq. 2 and Eq. 5, the tension-compression potential term can be expressed as

$$E_t(\mathbf{L}) = \sum_{(u,v) \in \mathcal{C}_t} T(L_u, L_v) = \sum_{(u,v) \in \mathcal{C}_t} \frac{1}{2} k_t (l'_{uv} - l_{uv})^2 \quad (9)$$

where $l_{uv} = 1$, and l'_{uv} is a function of both the coordinates and labels of nodes u, v

$$l'_{uv} = \sqrt{(x_u + X_u - x_v - X_v)^2 + (y_u + Y_u - y_v - Y_v)^2} \quad (10)$$

Bending potential. The bending potential (Eq. 3) of a spring (u, v, w) depends on the relationship of three nodes u, v, w . Let \mathcal{C}_b denote the set of cliques representing bending deformation. $(u, v, w) \in \mathcal{C}_b$, if nodes u, v, w are connected by a bending spring, that is $\theta_{uvw} = \pi$ and $l_{uv} = 1, l_{vw} = 1$. Using Eq. 4 and Eq. 5, the bending potential term can be expressed as

$$E_b(\mathbf{L}) = \sum_{\theta_{uvw} \in \mathcal{C}_b} B(L_u, L_v, L_w) = \sum_{\theta_{uvw} \in \mathcal{C}_b} \frac{1}{2} k_b (\theta'_{uvw} - \theta_{uvw})^2 \quad (11)$$

where $\theta_{uvw} = \pi$ and θ'_{uvw} can be computed by cosine theorem

$$\theta'_{uvw} = \arccos\left(\frac{l'_{uv}{}^2 + l'_{vw}{}^2 - l'_{uw}{}^2}{2l'_{uv}l'_{vw}}\right) \quad (12)$$

l'_{uv}, l'_{vw} and l'_{uw} can be calculated using the distance formula Eq. 10.

Now, we get the final structure of the 2D label MRF model $\mathcal{G} = (\mathcal{V}, \mathcal{C})$, where $\mathcal{C} = \mathcal{C}_t \cup \mathcal{C}_b$. The final energy

function, which consists of the data term (Eq. 8) and the deformation energy (Eq. 9 and Eq. 11), can be expressed as:

$$E(\mathbf{L}) = \underbrace{E_d(\mathbf{L})}_{\text{Data term}} + \underbrace{E_t(\mathbf{L})}_{T\text{-potential}} + \underbrace{E_b(\mathbf{L})}_{B\text{-potential}} \quad (13)$$

4. Optimization

Optimizing the energy $E(\mathbf{L})$ is a challenging problem. Ishikawa [14] proposes a graph cuts based algorithm to get the exact optimization of the MRF energy, which should satisfy convex priors. Unlike the multi-label MRF, each node in 2D label MRF takes a two dimensional label, hence $E(\mathbf{L})$ cannot satisfy the convex priors condition and it is unable to be optimized by the method [14]. Y.Boykov *et al.* [32] propose α -expansion algorithm which has been widely used. Although only an approximate solution can be obtained, it can be applied in n -dimensional label MRF. In this paper, we apply a two-stage method to optimize $E(\mathbf{L})$. We first use α -expansion algorithm [19,32] to obtain an approximate solution, and then adjust the location of elements according to the force balance to get a more precise matching result.

The α -expansion algorithm starts from an initial state and optimizes the objective by a series of iterative moves based on graph cuts. Each move randomly selects a 2D label $(X_\alpha, Y_\alpha) \in \mathcal{L}$, and allows any node v to either keep its current label (X_v, Y_v) or change to label (X_α, Y_α) by getting the min-cut of the corresponding graph. In the iteration, each move leads to a lower energy and the algorithm reaches convergence when no lower energy can be found.

However, the α -expansion algorithm cannot get precise solution and may lead to sharp deformations in some local positions. When the elastic object gets stable state after deformation, every element inside should be in equilibrium state. If the elements do not satisfy the equilibrium condition, they will move under the effect of resultant forces and the deformation energy will decrease during the move. Thus, we propose the location move based on force balance to adjust the location of each element to get smooth deformations.

As shown in Fig. 4, the location move starts from the initial state obtained by α -expansion. Each move randomly selects an element v to optimize its location. Let $F_{uv}^{\vec{}}$ be the force acting on spring (u, v) . The resultant force \vec{F}_v acting on v is the sum of components:

$$\vec{F}_v = \sum_{u, (u,v) \in \mathcal{C}_t} F_{uv}^{\vec{}} = \sum_{u, (u,v) \in \mathcal{C}_t} k_t \Delta \vec{l}_{uv} \quad (14)$$

where the vector $\Delta \vec{l}_{uv}$ denotes the deformation of (u, v) .

If the value of resultant force $|\vec{F}_v|$ is smaller than a chosen threshold T_f , the element is in balance. Otherwise, the

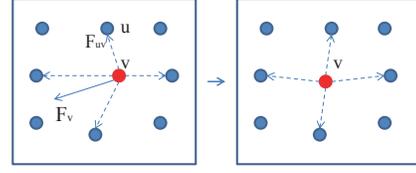


Figure 4. The location move algorithm. The left image is the initial state optimized by α -expansion. The right shows that node v move to the equilibrium position under the effect of resultant force.

element will move along the direction of \vec{F}_v . During the move, the element will reach a more stable state with the decrease of deformation energy. Thus, we can find the best location of element v along the direction of the resultant force by minimizing the data term and deformation energy:

$$L'_v = \arg \min_{L'_v} \{D(L'_v) + T(L'_v) + B(L'_v)\} \quad (15)$$

where $L'_v = (X'_v, Y'_v)$ is the new location after moving along \vec{F}_v . $T(L'_v) = \sum_{u, (u,v) \in \mathcal{C}_t} T_{uv}$ and $B(L'_v) = \sum_{(u,v,w) \in \mathcal{C}_b} B_{uvw}$.

The algorithm of the location move is shown in Alg. 1.

Algorithm 1 The algorithm of the location move

Initialization:

The initial label $\mathbf{L} = (\mathbf{X}, \mathbf{Y})$ optimized by α -expansion.
Threshold T_f .

Iteration:

- 1: **for** each node $v \in \mathcal{V}$ **do**
- 2: **for** each node $u, (u, v) \in \mathcal{C}_t$ **do**
- 3: Get the force $F_{uv}^{\vec{}} = k_t \Delta \vec{l}_{uv}$
- 4: **end for**
- 5: Get the resultant force $\vec{F}_v = \sum_{u, (u,v) \in \mathcal{C}_t} F_{uv}^{\vec{}}$
- 6: **if** $|\vec{F}_v| > T_f$ **then**
- 7: $L'_v = \arg \min_{L'_v} \{D(L'_v) + T(L'_v) + B(L'_v)\}$
- 8: **end if**
- 9: **end for**

Output:

- 10: Return label $L'_v = (X'_v, Y'_v)$ for each node.
-

5. Experiments and applications

Although many databases have been constructed for image matching, most of them focus on the problems of shape matching, such as MPEG-7 [21], silhouette database [27] and key point matching such as images from [25]. One of the main challenges in these databases is affine deformations which is not the emphasis of this paper. To the best of our knowledge, there is no database focusing on the deformable object matching problem with a quantized criterion, and providing the dense correspondence of all the pixels in real image pairs. As a result, many previous papers only

show several illustrative matching results without evaluation, and some papers just give the performance on handwriting or object recognition datasets. However, these experiments are not direct and effective to compare the matching performance of different models.

Therefore, to provide a quantitative benchmark for deformable object matching, we construct a database devoted for this purpose along with an evaluation criterion. To test the effectiveness of the proposed framework, we evaluate our approach on both the constructed deformation matching database and images from [22]. Meanwhile, we apply our method to handwriting recognition and compare it with other deformation models on United States Postal Service database (USPS) [1].

5.1. Deformation matching database

The constructed database contains 25 groups of deformable images. There is one image chosen to be the reference image and the remaining are target images in each group. The images in the database simulate various deformations of different elastic objects. They contain the deformation of human poses, human bodies doing sports, animals and some suitable examples selected from MPGE-7 [21]. Fig. 5 shows some images in the database. All the examples are binary images in our database, since it is difficult to provide the correspondence of all the pixels in color image. It will be much more convenient to provide a criterion to evaluate the matching performance on the binary images.

For a pair of binary images $I^{(q)}$, $q \in \{r, t\}$, there are two kinds of pixels in each image: object pixel set $\mathcal{P}_o^{(q)}$ and background pixel set $\mathcal{P}_b^{(q)}$. For each image $I^{(q)}$, there is an index matrix $M_j^{(q)}$, which represents whether the pixel j has a right matching pixel or not.

$$M_j^{(q)} = \begin{cases} 1, & \text{if } I_j^{(q)} = I_{\mathcal{M}_{(q)}(j)}^{(-q)} \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

where $\mathcal{M}_{(q)}(\cdot)$ is the mapping from image $I_{(q)}$ to $I_{(-q)}$. That is, the entry $M_j^{(q)}$ is 1 if the intensity of pixel $I_j^{(q)}$ is equal to the intensity of the corresponding pixel, where q refers to either the reference or target image.

Let $N_o^{(q)}$ be the number of pixels in $\mathcal{P}_o^{(q)}$ and $N_m^{(q)} = \sum_{j \in \mathcal{P}_o^{(q)}} M_j^{(q)}$ be the number of correct matching pixel of $\mathcal{P}_o^{(q)}$. The correct matching ratio P can be measured by the average of ratios in both \mathcal{P}_o^r and \mathcal{P}_o^t .

$$P = \frac{1}{2} \left(\frac{N_m^r}{N_o^r} + \frac{N_m^t}{N_o^t} \right) \quad (17)$$

We compare our method with related deformation models including *image distortion model* (IDM) [15], *two dimensional hidden Markov model* (HMM) [8, 30], *two dimensional hidden Markov distortion model* (HMDM) [16]

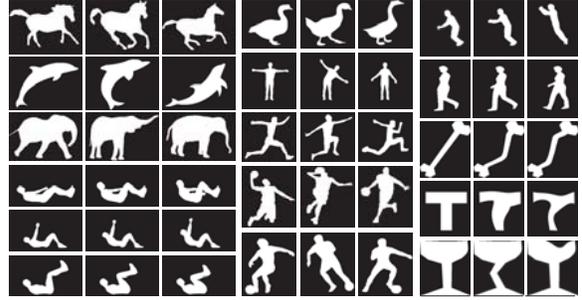


Figure 5. Some deformable examples in the constructed database, which is focusing on deformable object matching.

which is the combination of IDM and HMM, the approach of Shekhovtsov *et al.* [28] and Duchenne *et al.* [9]. Among the models, the first three are popular deformation models and have been successfully applied in handwriting recognition and face recognition. The last two models are most related MRF models. We do not compare with key point matching methods such as [6, 33] since they cannot get the dense correspondence of images.

To keep fair, we set the same parameters for all the methods. In the experiment, only the values of pixels are applied to measure the appearance similarity. Every pair of images is divided to $n \times 80$ or $80 \times n$ elements, where the ratio of $n/80$ is equal to the aspect ratio of the reference image. The maximum displacement is set as 20, and this means that we fix $\mathcal{L} = \{-20, \dots, 20\} \times \{-20, \dots, 20\}$ in our algorithm.

Tab. 1 shows the correct matching ratio of every group in the database. The proposed deformation model gets the best performance with 90.7% average correct matching ratio. Meanwhile, our model outperforms other methods on 22 groups in the database. The model [9] does not obtain a good matching performance as [28], because it models the matching problem in x and y direction independently. On the other hand, the optimization method of [9], which minimizes the energy alternately in x and y direction, easily lead to local minimum. IDM is the simplest deformation model and it gets satisfactory results in images with small deformations, such as comma and glass in the database. HMDM outperforms IDM and HMM on most groups. Fig. 6 shows the matching results of some examples by different deformation models. We can see that the distorted reference images of our model is most similar to the target images. Although there are various and complex deformations in the images, our method preserves the structure of distorted objects very well.

5.2. Deformable color image pairs

We test our model on color image pairs of the deformable database [22]. Since there are various and complex deformations in the objects depicted in the images, the database is very suitable for deformable object matching. Fig. 7 shows the matching results of three challenging and typical

| Methods | legbent | pushup | situp1 | situp2 | curvature | pullup | basketball | bone | comma | delphis | device | elephant | football |
|------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| IDM | 89.5 | 73.5 | 88.4 | 86.3 | 67.4 | 81.0 | 60.9 | 78.0 | 99.5 | 53.6 | 88.4 | 79.1 | 54.4 |
| HMM | 81.3 | 87.7 | 94.5 | 95.8 | 85.3 | 92.9 | 62.7 | 90.8 | 94.6 | 69.2 | 91.3 | 83.5 | 63.2 |
| HMDM | 83.8 | 88.5 | 95.0 | 96.4 | 86.8 | 89.3 | 66.3 | 91.4 | 95.5 | 70.9 | 93.4 | 85.7 | 65.2 |
| Shekhovtsov [28] | 88.5 | 94.8 | 93.3 | 92.9 | 87.4 | 93.8 | 73.9 | 94.4 | 97.5 | 83.3 | 97.2 | 91.2 | 77.5 |
| Duchenne [9] | 77.1 | 82.8 | 92.0 | 92.6 | 75.0 | 89.8 | 56.9 | 82.5 | 97.9 | 65.6 | 91.7 | 82.8 | 57.4 |
| Ours | 94.6 | 96.2 | 97.2 | 98.1 | 93.0 | 94.7 | 77.7 | 95.3 | 99.0 | 83.2 | 98.1 | 91.3 | 81.8 |

| Methods | glass | goose | horse | pose1 | pose2 | pose3 | jump | jumpjack | longjump | run | step | walk | Mean |
|------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| IDM | 95.1 | 68.3 | 65.9 | 67.6 | 62.8 | 67.7 | 47.4 | 67.8 | 54.6 | 61.7 | 62.0 | 81.2 | 71.9 |
| HMM | 84.5 | 77.0 | 74.2 | 84.6 | 76.3 | 82.9 | 73.8 | 75.5 | 78.1 | 63.8 | 87.6 | 76.1 | 81.1 |
| HMDM | 86.2 | 81.1 | 77.0 | 87.1 | 79.2 | 82.4 | 76.5 | 82.2 | 79.5 | 67.3 | 86.5 | 77.5 | 82.8 |
| Shekhovtsov [28] | 99.1 | 88.5 | 83.5 | 86.3 | 75.4 | 88.6 | 79.1 | 80.9 | 81.6 | 78.0 | 92.0 | 82.8 | 87.3 |
| Duchenne [9] | 97.5 | 72.9 | 72.6 | 74.1 | 68.2 | 74.9 | 49.5 | 68.9 | 62.9 | 64.4 | 70.6 | 57.4 | 75.2 |
| Ours | 98.2 | 92.9 | 86.7 | 88.8 | 84.7 | 91.6 | 82.3 | 87.8 | 88.1 | 84.3 | 95.3 | 87.8 | 90.7 |

Table 1. The average correct matching ratio (%) of every group in the deformation matching database.

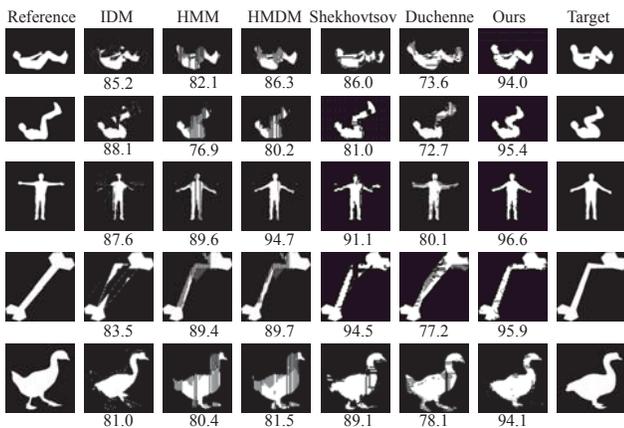


Figure 6. The matching results of some examples from the constructed database. The first column shows reference images and the last column shows target images with different deformations. The middle 6 columns show the distorted reference images by different models. The numbers under the images show the correct matching ratio(%).

image pairs. The first object in Fig. 7 is a surface with real non-rigid deformation and under different perspectives. The second and third image pairs are obtained with synthetic non-rigid deformations. In the experiments, we only use the RGB feature to measure the appearance similarity. We compare our model with the most related models [28] and [9]. As shown in Fig. 7, our model gets the most satisfactory matching results. In the second image pairs, both our model and the method [28] obtain promising matching results. In the first and third image pairs, the distorted reference images of our model are most similar to the target images. Thus, our model is effective to address the deformable object matching problem in complex deformations.

5.3. Handwriting recognition application

Handwriting recognition is one of the most important applications of deformation models. To test the effectiveness of our framework in applications, we apply the proposed model on handwriting recognition and compare with related

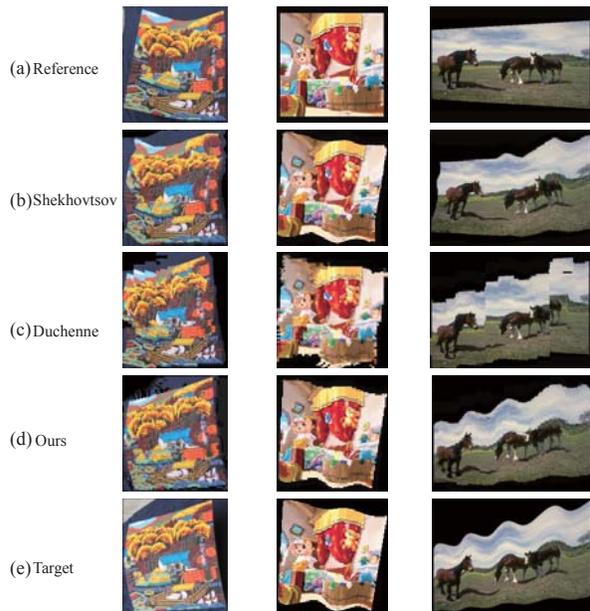


Figure 7. The matching results of deformable color image pairs. The reference images (a) are distorted to match the target images (e). (b), (c) and (d) show the distorted reference images by Shekhovtsov *et al.* [28], Duchenne *et al.* [9] and our model.

| Method | Error (%) |
|--------------------------|------------|
| Euclidean distance, 1-NN | 5.6 |
| SVM [26] | 4.0 |
| HMM [16] | 2.7 |
| IDM [16] | 2.4 |
| Hungarian matching [17] | 2.2 |
| 2D label MRF (Ours) | 1.9 |
| HMDM [16] | 1.9 |

Table 2. The performance of handwriting recognition on USPS database

deformation models on the well-known USPS database [1]. There are variable testing samples in the database, so it is very suitable to evaluate the proposed model. It contains 7291 training images and 2007 testing images. All the images are 16×16 pixels, and every pixel is one element in the

experiment. We fix the label set $\mathcal{L} = \{-5, \dots, 5\} \times \{-5, \dots, 5\}$, and apply pixel value and gradient feature to measure the similarity. We use 1-NN classifier to classify images. The results are shown in Tab. 2, and our model gets the best performance on USPS compared with related deformation models. The result shows that the proposed model can be effectively used in applications.

6. Conclusions and future work

In this paper, we have proposed a novel deformation decomposition model to represent various deformations and formulated the DDM as a 2D label MRF for deformable object matching. To get accurate matching performance, we have presented a two-stage optimization method containing graph cuts based α -expansion move and force balance based location move. Besides, we have proposed a generalized framework of 2D label MRF and extended the concept to n -dimensional label MRF. In the experiments, we have built a deformation matching database to provide a quantitative benchmark for deformable object matching. We have tested the proposed model on both the constructed database and color image pairs. Furthermore, we have applied our model on handwriting recognition. Experimental results show that our framework outperforms previous approaches.

Our framework for image matching can be readily extended to object classification. In the future, we will apply our model to object classification with the kernel method [31]. One of the problems in our model is that it cannot describe the 3D deformation, such as torsional deformation. We will extend our model to 3D label MRF and apply it on the 3D object matching problem.

Acknowledgement. This work is funded by the National Basic Research Program of China (Grant No. 2012CB316302), National Natural Science Foundation of China (Grant No. 61322209 and Grant No. 61175007), the National Key Technology R&D Program (Grant No. 2012BAH07B01).

References

- [1] United States Postal Service Handwritten Database. <http://ftp.kyb.tuebingen.mpg.de/pub/bs/data/>.
- [2] J. R. Barber. *Intermediate mechanics of materials*. New York:McGraw-Hill, 2001.
- [3] F. P. Beer, E. R. Johnston, and J. T. DeWolf. *Mechanics of materials*. McGraw-Hill Book Company, New York, 1992.
- [4] P. Bochev and R. Lehoucq. Energy principles and finite element methods for pure traction linear elasticity. *Comput. Methods Appl. Math.*, 11:173–191, 2011.
- [5] I. V. Chenchiah and P. D. Shipman. An energy-deformation decomposition for morphoelasticity. *arXiv preprint arXiv:1210.7174*, 2012.
- [6] M. Cho, J. Lee, and K. M. Lee. Feature correspondence and deformable object matching via agglomerative correspondence clustering. In *ICCV*, 2009.
- [7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.
- [8] D.S.Kuo and O.Agazzi. Keyword spotting in poorly printed documents using pseudo 2d hidden markov models. *IEEE TPAMI*, 16:842–848, 1994.
- [9] O. Duchenne, A. Joulin, and J. Ponce. A graph-matching kernel for object categorization. In *ICCV*, 2011.
- [10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE TPAMI*, 32:1627–1645, 2010.
- [11] M. Fischler and R. Elschlager. The representation and matching of pictorial structures. *IEEE Trans. on Computer*, 22:67–92, 1973.
- [12] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense image registration through mrfs and efficient linear programming. *Medical image analysis*, 12:731–741, 2008.
- [13] R. Grzeszczuk and D. Terzopoulos. Automated learning of muscle-actuated locomotion through control abstraction. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. ACM, 1995.
- [14] H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE TPAMI*, 25:1333–1336, 2003.
- [15] D. Keysers, J. Dahmen, T. Theiner, and H. Ney. Experiments with an extended tangent distance. In *ICPR*, 2000.
- [16] D. Keysers, T. Deselaers, C. Gollan, and H. Ney. Deformation models for image recognition. *IEEE TPAMI*, 29:1422–1435, 2007.
- [17] D. Keysers, T. Deselaers, and H. Ney. Pixel-to-pixel matching for image recognition using hungarian graph matching. *Pattern Recognition*, pages 154–162, 2004.
- [18] A. K.Jain, Y. Zhong, and S. Laksmanan. Object matching using deformable templates. *IEEE TPAMI*, 18:267–278, 1996.
- [19] V. Kolmogorov. What energy functions can be minimized via graph cuts?. *IEEE TPAMI*, 26:147–159, 2004.
- [20] L. Ladicky, P. Sturges, C. Russell, S. Sengupta, Y. Bastanlar, W. Clocksin, and P. H.S.Torr. Joint optimisation for object class segmentation and dense stereo reconstruction. In *BMVC*, 2010.
- [21] L. J. Latecki and R. Lakamper. Shape similarity measure based on correspondence of visual parts. *IEEE TPAMI*, 22:1–6, 2000.
- [22] H. Ling and D. W. Jacobs. Deformation invariant image matching. In *ICCV*, 2005.
- [23] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 2003.
- [24] T. McNerney and D. Terzopoulos. A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4d image analysis. *Computerized Medical Imaging and Graphics*, 19:69–83, 1995.
- [25] Mikołajczyk, Krystian, and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60:63–86, 2004.
- [26] B. Schölkopf. Support vector learning. 1997.
- [27] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE TPAMI*, 25:550–571, 2004.
- [28] A. Shekhovtsov, I. Kovtun, and V. Hlavac. Efficient mrf deformation model for non-rigid image matching. In *CVPR*, 2007.
- [29] C. Truesdell and W. Noll. *The non-linear field theories of mechanics*. Springer, 2004.
- [30] S. Uchida and H. Sakoe. A monotonic and continuous two-dimensional warping based on dynamic programming. In *ICPR*, 1998.
- [31] C. Wallraven, B. Caputo, and A. Graf. Recognition with local features: the kernel recipe. In *ICCV*, 2003.
- [32] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE TPAMI*, 23:1222–1239, 2001.
- [33] F. Zhou and F. D. la Torre. Deformable graph matching. In *CVPR*, 2013.