Circulant Dissimilarity Based Shape Registration for Object Segmentation

Xunxun Zeng[a], Fei Chen[a,*], Meiqing Wang[a], Choi-Hong Lai[b]
[a]College of Mathematics and Computer Science, Fuzhou University, China
[b]Department of Mathematical Sciences, University of Greenwich, UK

Abstract. A shape prior based object segmentation is developed in this paper by using a shape transformation distance to constrain object contour evolution. In the proposed algorithm, the transformation distance measures the dissimilarity between two unaligned shapes by cyclic shift, which is called "circulant dissimilarity". This dissimilarity with respect to translation and rotation of the object shape is represented by circular convolution, which could be efficiently computed by using fast Fourier transform. Given a set of training shapes, the kernel density estimate is adopted to model shape prior. By integrating low-level image feature, high-level shape prior and transformation distance, a variational segmentation model is proposed to solve the transformation invariance of shape prior. Numerical experiments demonstrate that circulant dissimilarity based shape registration outperforms the iterative optimization on explicit pose parameters, and show promising results and highlight the potential of the method for object registration and segmentation.

Keywords: Segmentation, Circulant dissimilarity, Shape prior, Kernel density estimation, Level set.

1 Introduction

The goal of object segmentation is to extract the object-of-interest $\mathbf{q} : \Omega \to \{0, 1\}$ from a given image $\mathbf{I} : \Omega \to \mathbb{R}$, where $\mathbf{q}$ is the object label on the image domain Ω. For any pixel $x \in \Omega$, $\mathbf{q}(x) = 1$ if x is an object pixel; 0 otherwise. In the case of single-object segmentation, $\mathbf{q}$ is commonly known as an object shape. As a classical and fundamental problem in computer vision, object segmentation has been widely studied.[1–4] From a probability based perspective, object segmentation could be treated as a Bayesian posterior estimation problem: $p(\mathbf{q}|\mathbf{I}) = p(\mathbf{I}|\mathbf{q})p(\mathbf{q})/p(\mathbf{I})$, where $p(\mathbf{I})$ is a constant once $\mathbf{I}$ is given. In terms of logarithmic likelihood $E = -\ln p$, we can minimize the posterior energy:

$$E(\mathbf{q}) = -\log p(\mathbf{I}|\mathbf{q}) - \log p(\mathbf{q}) \tag{1}$$

1

Assuming that the independence of pixels in the image, the data model of $p(\mathbf{I}|\mathbf{q})$ can be written as

$$E_d(\mathbf{q}) = -\log p(\mathbf{I}|\mathbf{q}) = -\int_\Omega \mathbf{q}(x) \log p_{in}(\mathbf{I}(x)) + (1 - \mathbf{q}(x)) \log p_{out}(\mathbf{I}(\mathbf{x}))dx, \qquad (2)$$

where $p_{in}(x)$ and $p_{out}(x)$, for every pixel $x \in \Omega$, the probabilities that this pixel is inside the object and background, respectively. The distribution of probability $p(\mathbf{q})$ is called shape prior model since it specifies a prior bias among the desired object and is independent of data observation. The classic active shape model (ASM)[5] encodes statistical prior of object shape for segmentation. The shape is represented as a set of landmark points, and the variation of shape is constrained by the point distribution model which is inferred from a training set of shapes. Based on the seminal work of level set method,[6] shape prior based variational approaches have gained significant attention in object segmentation. Many statistical and learning models of shape priors have been proposed, such as Gaussian distribution[7] , kernel density estimation,[8,9] sparse representation,[10] manifold learning,[11] deep learning,[12] etc. Kernel density estimation (KDE)[9] is quite suitable for segmenting objects of a known class in the image according to their possible similar shapes. Given a set of aligned training shapes $\{\mathbf{p}_1, \cdots, \mathbf{p}_N\}$, the shape distribution can be modeled by the kernel density estimate:

$$p(\mathbf{q}) = \frac{1}{N\sigma} \sum_{i=1}^{N} \kappa_\sigma \left(\frac{\mathbf{q} - \mathbf{p}_i}{\sigma} \right) \qquad (3)$$

where σ is the width of the Gaussian kernel function, and $\kappa_\sigma(\mathbf{q}) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{\mathbf{q}^2}{2})$.

However, the segmented object often has similar shapes in different poses in real applications. An important problem of transformation invariance arises from the requirement of these shape driven object segmentations. Considering the transformations of translation h and rotation θ, the

Bayesian inference problem is therefore rewritten as

$$p(\mathbf{q}, h, \theta | \mathbf{I}) = p(\mathbf{I} | \mathbf{q}, h, \theta) p(\mathbf{q}, h, \theta) / p(\mathbf{I}) \tag{4}$$

Furthermore, a uniform prior is commonly assumed with respect to h and θ, and $\mathbf{q}$ is independent of these parameters, i.e. $p(\mathbf{q}, h, \theta) = p(\mathbf{q})$. Early ideas of handling this issue[13–15] was to introduce a set of explicit pose parameters to describe the shape transformation, i.e. $\mathbf{q}_{h,\theta}(x) = \mathbf{q}(R_\theta x + h)$, which is iteratively optimized the best transformation parameters by an alternating minimization procedure. However, such methods are difficult to balance the time step size in gradient descent. Cremers et al. proposed an intrinsic alignment[8] to deal with the pose issue. However, the training shapes requires normalization with respect to translation, scale and rotation in advance. In numerical experiments, it needs to accurately compute the center of mass and the principal axes of the shape for alignment. Based on sparse representation,[10,16,17] invariance of shape prior can be implemented by searching a set of transformations aiming at the error to be sparse between the transformed test shape and linear combination of training shapes. However, this approach is time consuming and not suitable in low dimensional space.

In this paper, a new shape registration method is introduced for object segmentation by exploiting shape circulant dissimilarity. It is observed that the object shape can be approximated by the training shapes after a certain cyclic shift. This circulant dissimilarity is used as a shape distance of transformation between the evolutionary shape and the reference shape, and could be easily integrated into variational segmentation model. By using circulant shift, the proposed shape registration becomes very simple and relies solely on the kernel circulant matrix. Moreover, the computational cost can be much reduced by fast Fourier transform. In contrast to existing

3

Fig 1 The partly occluded shape can be approximated by a sparse linear combination of horizontal cyclic shifted shapes.

approaches for transformation invariance in the level set framework, the proposed closed-form solution removes the need to iteratively optimize explicit pose parameters.

2 Shape Transformation Distance via Circulant Dissimilarity

In order to define a transformation distance or dissimilarity measure for two given shapes, we introduce a probabilistic definition of shape.[18] A shape on an image domain Ω is defined as a function $\mathbf{q} : \Omega \to [0, 1]$, which assigns to each pixel $x \in \Omega$ a probability $\mathbf{q}(x)$ that x is part of the object. By selecting $\tau \in [0, 1]$, it is easy to get the traditional binary shape of the object $(\mathbf{q})_\tau = \{x | \mathbf{q}(x) \geq \tau\}$ and the background of image $(\mathbf{q})_\tau^C = 1 - (\mathbf{q})_\tau$.

2.1 Translation Invariance via Cyclic Shift

In this section, we stick to definitions for 1D signals for simplicity of presentation. The extension to 2D is straightforward. Denote by $\mathbf{q}$ the observed object shape, $\mathbf{q} = [q_0, \cdots, q_{n-1}]^\top$, and we define a shift operator $T^l : T^l(q_0, \cdots, q_{n-l-1}, q_{n-l}, \cdots, q_{n-1}) = (q_{n-l}, \cdots, q_{n-1}, q_0, \cdots, q_{n-l-1})$, where l is the number of shifted elements to the right, and we have $T^0 = T$, $T^{n+l} = T^l$. Given a reference shape $\mathbf{p}$, the object shape $\mathbf{q}$ can be approximated by a sparse linear combination of circulant shifted $\mathbf{p}$ (see Fig. 1). This can take the following form,

$$\mathbf{q} \approx \sum_{i=0}^{n-1} \kappa_h(\mathbf{q}, T^i \mathbf{p}) T^i \mathbf{p} \tag{5}$$

4

where the weight function $\kappa_h(\cdot)$ is used to measure the similarity between the shape $\mathbf{q}$ and cyclic shifted shape $T^i\mathbf{p}$. If the weight function κ_h is chosen as a Gaussian kernel, we have $\kappa_h(T^i\mathbf{q}, T^j\mathbf{p}) = \kappa_h(\mathbf{q}, T^{j-i}\mathbf{p})$. By cyclic shift, the above equation can also be written in matrix notation

$$
\begin{bmatrix} q_0 \\ q_1 \\ \vdots \\ q_{n-1} \end{bmatrix} \approx \begin{bmatrix} p_0 & p_{n-1} & \cdots & p_1 \\ p_1 & p_0 & \cdots & p_2 \\ \vdots & \vdots & \ddots & \vdots \\ p_{n-1} & p_{n-2} & \cdots & p_0 \end{bmatrix} \begin{bmatrix} \kappa_h(\mathbf{q}, T^0\mathbf{p}) \\ \kappa_h(\mathbf{q}, T^1\mathbf{p}) \\ \vdots \\ \kappa_h(\mathbf{q}, T^{n-1}\mathbf{p}) \end{bmatrix}
$$

$$
= \begin{bmatrix} \kappa_h(\mathbf{q}, T^0\mathbf{p}) & \kappa_h(\mathbf{q}, T^{n-1}\mathbf{p}) & \cdots & \kappa_h(\mathbf{q}, T^1\mathbf{p}) \\ \kappa_h(\mathbf{q}, T^1\mathbf{p}) & \kappa_h(\mathbf{q}, T^0\mathbf{p}) & \cdots & \kappa_h(\mathbf{q}, T^{n-2}\mathbf{p}) \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_h(\mathbf{q}, T^{n-1}\mathbf{p}) & \kappa_h(\mathbf{q}, T^{n-2}\mathbf{p}) & \cdots & \kappa_h(\mathbf{q}, T^0\mathbf{p}) \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_{n-1} \end{bmatrix},
$$

i.e.

$$
\mathbf{q} \approx K\mathbf{p} \tag{6}
$$

where

$$
K = \begin{bmatrix} \kappa_h(\mathbf{q}, T^0\mathbf{p}) & \cdots & \kappa_h(\mathbf{q}, T^1\mathbf{p}) \\ \vdots & \ddots & \vdots \\ \kappa_h(\mathbf{q}, T^{n-1}\mathbf{p}) & \cdots & \kappa_h(\mathbf{q}, T^0\mathbf{p}) \end{bmatrix}. \tag{7}
$$

From the above equation, it can be seen that the shape similarity between $\mathbf{q}$ and $\mathbf{p}$ is dependent on the kernel matrix K, and K is a circulant matrix.[19] Based on the theory of circulant matrices,[20] K is often denoted by $K = C(\mathbf{k}_h)$, where $\mathbf{k}_h = [\kappa_h(\mathbf{q}, T^0\mathbf{p}), \cdots, \kappa_h(\mathbf{q}, T^{n-1}\mathbf{p})]^T$ is the cyclic element. Note that if $\mathbf{p}$ is 2D shape, then $\mathbf{k}_h$ is also 2D kernel matrix. When $\kappa_h(\mathbf{q}, T^i\mathbf{p}) \to 1$,

(a) Test shape 'A' (b) Polar coordinates (c) Cyclic shift along θ axes (d) Cartesian coordinates
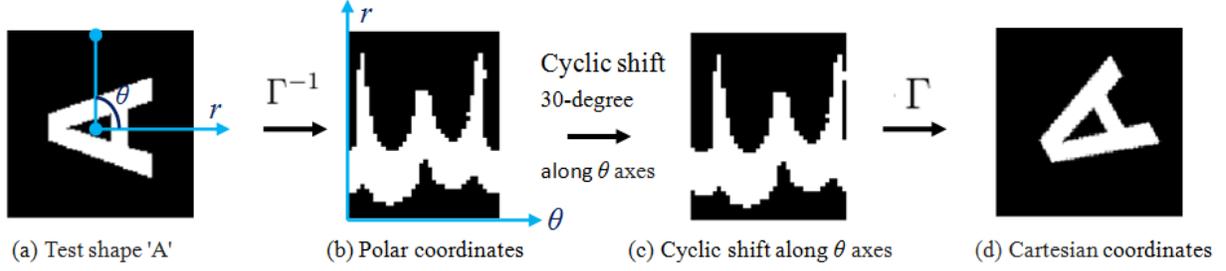
Fig 2 Rotation invariance by cyclic shift in polar coordinates. A test shape (a) is converted from Cartesian coordinates to polar coordinates (b). By horizontal cyclic shift 30-degree along θ axes (c), a rotated shape (d) is obtained by using conversion from polar coordinates to Cartesian coordinates.

it means that $T^i\mathbf{p}$ is very similar to $\mathbf{q}$. In particular, $\kappa_h(\mathbf{q}, T^0\mathbf{p}) = 1$ means $T^0\mathbf{p} = \mathbf{q}$. On the contrary, if $\kappa_h(\mathbf{q}, T^i\mathbf{p}) \to 0$, then $T^i\mathbf{p}$ is very different from $\mathbf{q}$. Since our problem focuses on translation transformation, we apply a simple method to locate object shape accurately by setting the largest element in $\mathbf{k}_h$ to be one and others zero (see Fig. 1). Since K is a circulant matrix, we have the circular convolution

$$\mathbf{q} \approx C(\mathbf{k}_h)\mathbf{p} = \mathbf{k}_h \star \mathbf{p} \qquad (8)$$

Given two shapes $\mathbf{q}$ and $\mathbf{p}$, a simple measure of their dissimilarity with respect to translation transformation is given by their L_2-distance in Ω:

$$Dis(\mathbf{q}, \mathbf{p}) = \min_{\mathbf{k}_h} \int_{\Omega} (\mathbf{q} - \mathbf{k}_h \star \mathbf{p})^2 \, dx \qquad (9)$$

While translation is usually written in Cartesian coordinates, rotation and scaling are simpler in polar coordinates. The above approach may be extended to a shape distance which is invariant to rotation and scaling.

2.2 Rotation Invariance via Cyclic Shift

The invariance for the cases of translation and rotation is detailed here. When a shape $\mathbf{q}$ is given in Cartesian coordinates, it can be converted to polar coordinates by a transformation Γ^{-1}, such that $\mathbf{q} = \Gamma(\Gamma^{-1}(\mathbf{q}))$, where Γ from polar to rectangular coordinates is defined by $\Gamma(r, \theta) = (r\cos\theta, r\sin\theta) = (x, y)$, and Γ^{-1} is its inverse transformation. Note that rotation invariance could be achieved by cyclic shift on polar axes. Extensions to scale is similar but is not discussed here. For example, a test shape 'A' is transformed from Cartesian coordinates to polar coordinates (see Fig. 2). By cyclic shifting 30-degree along θ axes, a rotated shape could be obtained by conversion from polar coordinates to Cartesian coordinates. Like translation invariance on shapes $\mathbf{p}$ and $\mathbf{q}$, we can perform cyclic shift along θ coordinate, and obtain a kernel matrix $\mathbf{k}_\theta$, and have

$$\Gamma^{-1}(\mathbf{q}) \approx \mathbf{k}_\theta \star \Gamma^{-1}(\mathbf{p}) \tag{10}$$

Assuming that translation takes place before rotation, the transformation dissimilarity between two shapes $\mathbf{p}$ and $\mathbf{q}$ is proposed by considering translation and rotation,

$$Dis(\mathbf{q}, \mathbf{p}) = \min_{\mathbf{k}_h, \mathbf{k}_\theta} \int_\Omega \left(\mathbf{q} - \Gamma(\mathbf{k}_\theta \star \Gamma^{-1}(\mathbf{k}_h \star \mathbf{p})) \right)^2 dx \tag{11}$$

3 Object Segmentation via Shape Registration

3.1 Energy Formulation

Considering a set of aligned training shapes $\chi = \{\mathbf{p}_1, \cdots, \mathbf{p}_N\}$, the kernel density estimate is used to model the shape distribution. By combining image data term (2), shape prior term (3), and the proposed transformation term (10), the energy function for object segmentation can be formulated
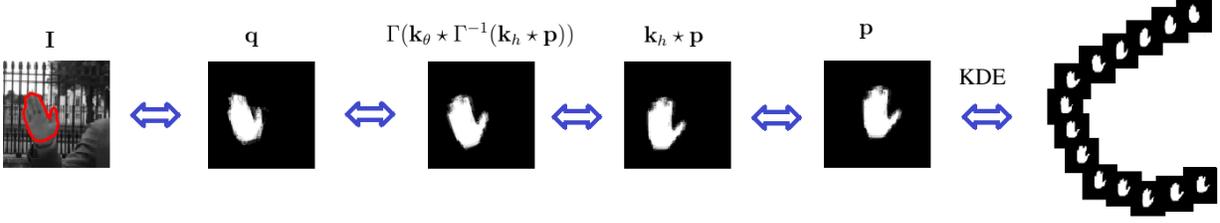
Fig 3 The shape registration (middle) is proposed to link low-level image data (left) with high-level shape prior (right) for object segmentation.

as

$$E(\mathbf{q}, \mathbf{p}, \mathbf{k}_h, \mathbf{k}_\theta) = -\int_\Omega \mathbf{q}(x)\mathbf{e}(x)dx + \lambda_1 \int_\Omega \left(\mathbf{q}(x) - \Gamma(\mathbf{k}_\theta \star \Gamma^{-1}(\mathbf{k}_h \star \mathbf{p}(x)))\right)^2 dx$$

$$-\lambda_2 \log\left(\frac{1}{N\sigma}\sum_{i=1}^N \kappa_\sigma\left(\frac{\mathbf{p} - \mathbf{p}_i}{\sigma}\right)\right). \tag{12}$$

Here, λ_1 and λ_2 are positive constants. The first term on ther right is the simplified form of (2), and $\mathbf{e}(x) = \log\left(p_{in}(\mathbf{I})/p_{out}(\mathbf{I})\right)$. The sencond term adds an additional force aiming at maxmizing the transformation similarity between the evolutionary shape $\mathbf{q}$ and the reference shape $\mathbf{p}$. The last term enforces $\mathbf{p}$ to be the estimation shape inferred from the training set χ. When the prior shape is only a given shape, the last shape statistical term can be neglected. If the transformed shape $\mathbf{p}$ is expected as the desired shape for extraction, we can define $\mathbf{q} = \Gamma(\mathbf{k}_\theta \star \Gamma^{-1}(\mathbf{k}_h \star \mathbf{p}))$. Fig. 3 illustrates the effect of shape registration, which couples the shape-based cue and intensity-based cue to establish a correspondence between them. By alternating minimization, registration and segmentation are carried out simultaneously.

There are three obvious advantages of the proposed object segmentation by using transformation constraint. First, the proposed transformation term is consistent with the shape probabilistic representation, and can be easily integrated into data-driven variational frame for object segmentation. Second, shape alignment is obtained by cirlulant dissimilarity, which quickly finds the most

similar shape by incorporating information from all transformed shapes. Third, a closed-form and fast solution can be derived for circulant similarity, and it removes the need to iteratively optimize explicit pose parameters.

3.2 Model in Low Dimensional Representation

Since shape is defined by probability, the space of shapes is convex so that principal component analysis (PCA) can be used to reduce dimensions of shape data. The shape space of χ spanned by the first $m \leq N$ eigenmodes $\{\psi_1, \cdots, \psi_m\}$ can be written as $\chi_m = \{\mathbf{p}_\alpha = \mu + \sum_{i=1}^{m} \alpha'_i \psi_i | \alpha'_i \in \mathbb{R}\}$ by PCA. Therefore, an arbitrary shape $\mathbf{p}$ can be approximated by a shape vector of the form $\alpha_{\mathbf{p}} = \Psi^T(\mathbf{p}_\alpha - \mu)$, where $\Psi = [\psi_1, \cdots, \psi_m]$ and $\alpha \in \mathbb{R}^{m \times 1}$. By PCA, the training set of shapes $\mathbf{p}_1, \cdots, \mathbf{p}_N$ can be reduced to a sequence of low dimensional coefficient vectors $\alpha_1, \cdots, \alpha_N$. By neglecting the constant terms, the variational model (12) in low-dimensional representation can be described by

$$E(\mathbf{q}, \alpha, \mathbf{k}_h, \mathbf{k}_\theta) = -\langle \mathbf{q}, \mathbf{e} \rangle + \lambda_1 \|\mathbf{q} - \Gamma(\mathbf{k}_\theta \star \Gamma^{-1}(\mathbf{k}_h \star \mathbf{p}_\alpha))\|^2$$

$$-\lambda_2 \log \left(\sum_{i=1}^{N} \kappa_\sigma \left(\frac{\alpha - \alpha_i}{\sigma} \right) \right). \tag{13}$$

To this end, we propose a shape registration based statistical shape prior which combines the efficiency of low-dimensional PCA-based methods.

4 Energy Minimization

When the PCA parameters $\{\Psi, \mu\}$ are known, the proposed model in Eq. (13) has three kinds of unknowns: $\mathbf{q}$, α, and the transformation kernel matrices $\{\mathbf{k}_h, \mathbf{k}_\theta\}$. To solve the model, we employ the alternating minimization algorithm by iteratively performing the following three steps: (i) updating $\mathbf{q}$ given α, $\mathbf{k}_h$ and $\mathbf{k}_\theta$, (ii) updating $\mathbf{k}_h$ and $\mathbf{k}_\theta$ given $\mathbf{q}$ and α, and (iii) updating α given $\mathbf{q}$, $\mathbf{k}_h$ and $\mathbf{k}_\theta$.

4.1 Updating $\mathbf{q}$

Given the estimate of $\mathbf{p}^{(t)} = \mu + \Psi \alpha^{(t)}$, $\mathbf{k}_h^{(t)}$ and $\mathbf{k}_\theta^{(t)}$, the subproblem on $\mathbf{q}$ can be formulated as

$$\min_{\mathbf{q}} \lambda_1 \|\mathbf{q} - \Gamma(\mathbf{k}_\theta^{(t)} \star \Gamma^{-1}(\mathbf{k}_h^{(t)} \star \mathbf{p}^{(t)}))\|^2 - \langle \mathbf{q}, \mathbf{e} \rangle. \tag{14}$$

The above $\mathbf{q}$-subproblem is convex, and its closed-form solution can be obtained by

$$\mathbf{q}^{(t)} = \Gamma(\mathbf{k}_\theta^{(t)} \star \Gamma^{-1}(\mathbf{k}_h^{(t)} \star \mathbf{p}^{(t)})) + \frac{1}{2\lambda_1} \mathbf{e}. \tag{15}$$

4.2 Updating $\mathbf{k}_h$ and $\mathbf{k}_\theta$

Given the estimate of the latent shape $\mathbf{q}^{(t)}$ and $\mathbf{p}^{(t)}$, the subproblem on $\mathbf{k}_h$ and $\mathbf{k}_\theta$ can be formulated as

$$\min_{\mathbf{k}_h, \mathbf{k}_\theta} \|\mathbf{q}^{(t)} - \Gamma(\mathbf{k}_\theta \star \Gamma^{-1}(\mathbf{k}_h \star \mathbf{p}^{(t)}))\|^2. \tag{16}$$

Assuming that translation takes place before rotation, the model above can be decomposed into

two subproblems,

$$
\begin{cases}
\mathbf{k}_h^{(t+1)} = \arg\min_{\mathbf{k}_h} \|\mathbf{q}^{(t)} - \mathbf{k}_h \star \mathbf{p}^{(t)}\|^2, \\
\mathbf{k}_\theta^{(t+1)} = \arg\min_{\mathbf{k}_\theta} \|\mathbf{q}^{(t)} - \Gamma(\mathbf{k}_\theta \star \Gamma^{-1}(\mathbf{k}_h^{(t+1)} \star \mathbf{p}^{(t)}))\|^2.
\end{cases}
\tag{17}
$$

Since accurate kernel matrixes are difficult to directly obtain from the above equations, so we use Gaussian kernel function to approximate the desired solution. As discussed in section 2.1, $\mathbf{k}_h$ measures the similarity between the patch $\mathbf{q}$ and $\mathbf{p}$ by cyclic shift. If σ_0 is the width of the chosen Gaussian kernel function, we have kernel matrix

$$
\tilde{\mathbf{k}}_h = \exp\left\{-\frac{1}{\sigma_0^2}\left(\|\mathbf{q}^{(t)}\|^2 + \|\mathbf{p}^{(t)}\|^2 - 2\mathbf{q}^{(t)} \star \mathbf{p}^{(t)}\right)\right\},
\tag{18}
$$

To accurately locate the object shape, we define $B_{max}(\mathbf{x})$ as threshold function that sets the largest element in $\mathbf{x}$ to be one and others zero. The closed-form solution to $\mathbf{k}_h$-subproblem can be obtained by

$$
\mathbf{k}_h^{(t+1)} = B_{max}(\tilde{\mathbf{k}}_h).
\tag{19}
$$

Similarly, we have

$$
\mathbf{k}_\theta^{(t+1)} = B_{max}\left\{\exp\left(-\frac{1}{\sigma_0^2}\left(\|\Gamma_{\mathbf{q}}^{-1}\|^2 + \|\Gamma_{\mathbf{p}}^{-1}\|^2 - 2\Gamma_{\mathbf{q}}^{-1} \star \Gamma_{\mathbf{p}}^{-1}\right)\right)\right\},
\tag{20}
$$

where $\Gamma_{\mathbf{q}}^{-1} = \Gamma^{-1}(\mathbf{q}^{(t)})$, and $\Gamma_{\mathbf{p}}^{-1} = \Gamma^{-1}(\mathbf{k}_h^{(t)} \star \mathbf{p}^{(t)})$.

11

4.3 Updating α

Given the shape $\mathbf{q}^{(t)}$, and kernel matrixes $\mathbf{k}_h^{(t+1)}$ and $\mathbf{k}_\theta^{(t+1)}$, the subproblem on α can be formulated as

$$\min_\alpha \|\mathbf{q}^{(t)} - \Gamma(\mathbf{k}_\theta^{(t+1)} \star \Gamma^{-1}(\mathbf{k}_h^{(t+1)} \star \mathbf{p}_\alpha))\|^2 - \lambda_2 \log\left(\sum_{i=1}^N \kappa_\sigma\left(\frac{\alpha - \alpha_i}{\sigma}\right)\right). \qquad (21)$$

With $\mathbf{p}_\alpha = \mu + \Psi\alpha$, the α-subproblem is non-convex and can be solved by using gradient descent:

$$\alpha^{(t+1)} = \alpha^{(t)} + \Delta t\left\{\left\langle 2D, \Gamma(\mathbf{k}_\theta^{(t+1)} \star \Gamma^{-1}(\mathbf{k}_h^{(t+1)} \star \Psi))\right\rangle + \frac{\lambda_2}{\sigma^2}\frac{\sum_{i=1}^N (\alpha^{(t)} - \alpha_i)\mathcal{K}_i}{\sum_{i=1}^N \mathcal{K}_i}\right\}, \qquad (22)$$

where $D = \mathbf{q}^{(t)} - \Gamma(\mathbf{k}_\theta^{(t+1)} \star \Gamma^{-1}(\mathbf{k}_h^{(t+1)} \star \mathbf{p}_{\alpha^{(t)}}))$, $\mathcal{K}_i = \kappa_\sigma(\frac{\alpha^{(t)} - \alpha_i}{\sigma})$, and Δt is the time step size.

Algorithm 1 Circulant similarity based shape registration for object segmentation

---

Input: Test image I, learned PCA parameters $\{\Psi, \mu\}$, and low-dimensional shape vectors $\alpha_1, \cdots, \alpha_N$.
Output:Plot of $(\hat{\mathbf{q}})_\tau$ for segmentation.
Initialization: Choose appropriate $\lambda_1, \lambda_2, \tau, \sigma_0$, and σ for kernel density estimation, and initialize $\alpha^{(0)}, \mathbf{k}_h^{(0)}$ and $\mathbf{k}_\theta^{(0)}$.
for $t = 1 : MaxIter$ **do**
    Find the region $(\mathbf{p}^{(t)})_\tau$, where $\mathbf{p}^{(t)} = \mu + \Psi\alpha^{(t)}$;
    Compute $\mathbf{e}(x) = \log(\frac{p_{in}(I)}{p_{out}(I)})$ for each $x \in \Omega$;
    Estimate the object region $\mathbf{q}^{(t)}$ according to (15);
    Compute the $\mathbf{k}_h^{(t+1)}$ according to (19);
    Compute the $\mathbf{k}_\theta^{(t+1)}$ according to (20);
    Update $\alpha^{(t+1)}$ by (22);
    if $\|\mathbf{q}^{(t)} - \mathbf{q}^{(t-1)}\|/\|\mathbf{q}^{(t)}\| \leq \epsilon$ & $\|\alpha^{(t+1)} - \alpha^{(t)}\|/\|\alpha^{(t)}\| \leq \epsilon$ **then**
        break
    end if
end for

Finally, Algorithm 1 summarizes the main steps of our alternating minimization optimization. The algorithm involves multiple circular convolution, which can be computed by FFT.[19] For a shape with n pixels, the three steps in the optimization procedure have the complexity of

$O(n \log_2 n)$, which makes this new algorithm very efficient for object segmentation.

5 Experimental Results

This section evaluates the efficiency and robustness of the proposed method on two shape datasets: walking-person dataset and our own hand-posture dataset. Three test sequences were carried out to demonstrate transformation invariance, covering the case of different types of images including partial occlusions, deformation, background clutter, etc. In our experiments, the initial translation parameter in the first frame was given manually. When moving on to a new frame, the final translation parameter of previous frame was used as the initial translation estimate for the segmentation of the current frame. We set $MaxIter = 30$, $\lambda_1 = 10$, $\tau = 0.5$, and $\Delta t = 0.1$. The parameter α was initialized as $[0, \cdots, 0]^T$. The involved parameters λ_2, σ and σ_0 in the proposed algorithm are fixed to 1.

5.1 Track a Walking Person

In order to test our model that was capable to take account of shape prior and tranformation invariance, the proposed model was applied to segment a partially occluded walking person. The dataset is based on,[18] which has 151 training shapes from a consecutive sequence. Fig. 4 shows five samples from the walking sequence. Due to partial occlusion, shape model[8] which used explicit pose parameters could not provide reliable location for shape prior in the first sample, thus yielding a dissatisfactory segmentation. It could be easily observed that the proposed method was more accurate because the closed form solution to cyclic shift removes the need to iteratively update local estimates of explicit pose parameters.
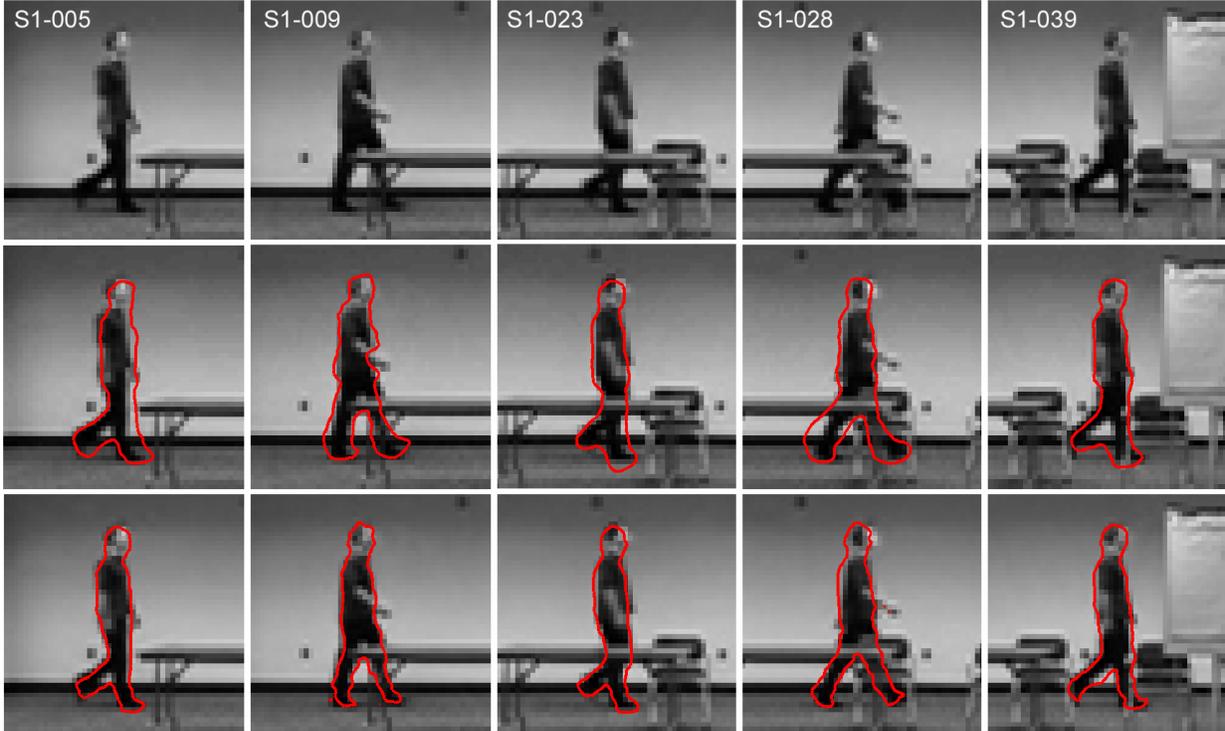
Fig 4 Typical segmentation comparison on tracking a walking person. The frames 5, 9, 23, 28, and 39 are shown in the first row. Segmentation results by iterative optimization on explicit pose parameters[8] are given in the middle row. Our segmentation results are presented in the last row.

(a)   (b)   (c)   (d)
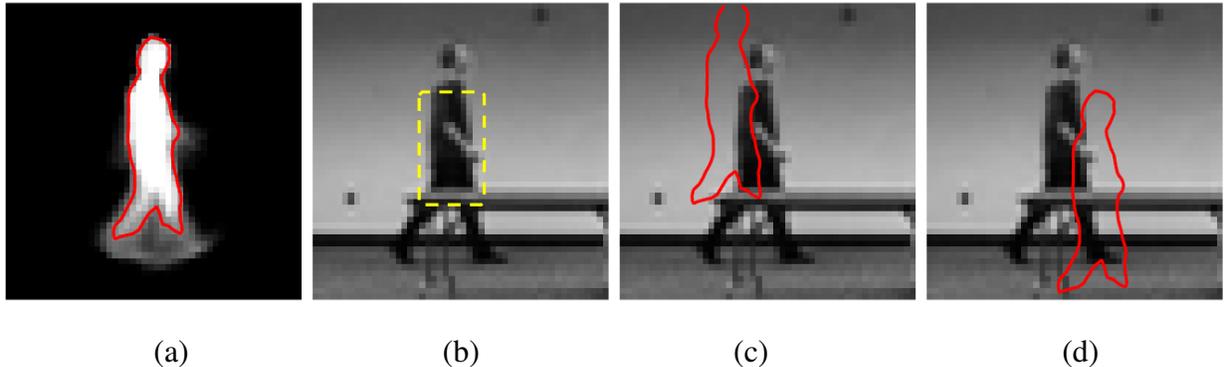
Fig 5 Robust initialization. (a) The mean shape of all training shapes and its boundary (0.5-level set). (b) The different locations of initial curves were randomly chosen within a local range (yellow box). Two typical initial localtion were shown in (c) and (d). Noted that the initial locations were a certain distance from the walking person.

The proposed approach is more robust to the location of initial evolution curve than the existing iterative optimization methods. For comparison, the evolution shapes were simply initialized as the mean shape of all training shapes. The red boundary was the 0.5-level set of the mean shape, as shown in Fig. 5(a). In order to validate the robust initialization, 50 different locations of initial

14
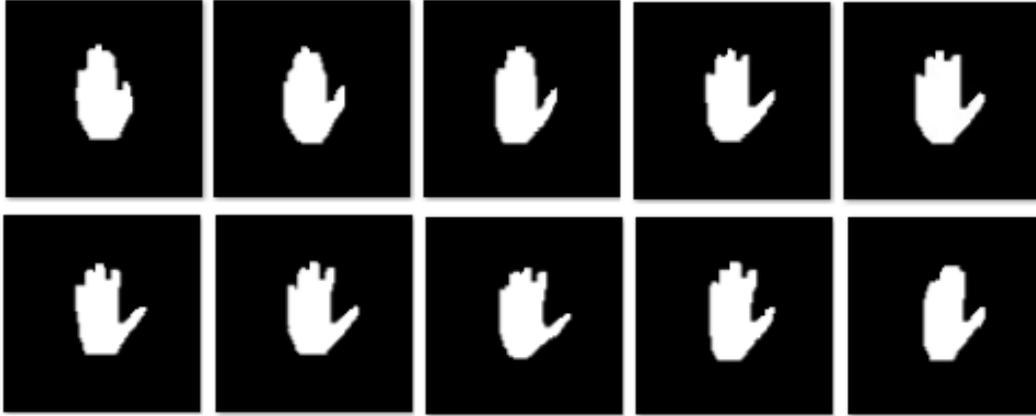
Fig 6 Selected sample shapes from the hand-posture dataset.

curves were randomly chosen within a local range (yellow box on Fig. 5(b)). Since the location of initial evolution curve was sensitive to active contours based models, the method[8] only had a 82 percent success rate in segmenting the walking person. Two typical failures of initial localtion were shown in Fig. 5(c) and 5(d). Noted that the initial locations were a certain distance from the walking person. However, the proposed method adopts circulant dissimilarity to overcome these drawbacks, and achieve a success rate of 100% under the same conditions.

5.2 Hand Posture Segmentation

In order to demonstrate the effectiveness and robustness for transformation invariance, two hand-posture image sequences with varying positions of the hands and complex background were used to examine the performance of the proposed model. Our own training dataset for this example consists of 100 binary shapes with 100×100 pixels. Fig. 6 shows some of the shapes from the dataset. For the two experiments, we kept all involved parameters constant. Due to circulant dissimilarity to translation and rotation, there was no need to iteratively optimize explicit pose parameters to align the evolutionary shape with the reference shape. The final segmentation is shown in the bottom row of Fig. 7. Notice that the red contour accurately outlines the hand skin surface at

Fig 7 Typical comparison on hand-posture sequence 1. The original frames 3, 17, 22, 46, and 56 are shown in the first row. Segmentation results by iterative optimization on explicit pose parameters[9] are given in the second row. Segmentation results by exhaustive pose parameters search[12] are given in the third row. Our segmentation results are presented in the last row.

different locations and rotations. From Fig. 7 and Fig. 8, we can see the proposed model with circulant dissimilarity makes the segmentation process robust to deformations and bad contrast. Compared with the state-of-the-art shape prior-based method, the proposed method is competitive in segmentation results but is much more efficient. As shown in Table 1, Dice Coefficient (DC) and Hausdorff Distance (HD)[21] were used to quantitatively evaluate the segmentation results. It is clear that the proposed method significantly outperforms the iterative optimization based pose estimation.[9] Compared with the exhaustive search,[12] our method is more than 6 times faster even

16

Fig 8 Typical comparison on hand-posture sequence 2. The original frames 3, 11, 21, 30, and 40 are shown in the first row. Segmentation results by iterative optimization on explicit pose parameters[9] are given in the second row. Segmentation results by exhaustive pose parameters search[12] are given in the third row. Our segmentation results are presented in the last row.

Table 1 Comparisons on hand-posture dataset by using mean Dice coefficient, mean running time (seconds) of all frames in each image sequence.

Method	Sequence 1			Sequence 2		
	DC (%)	HD (px)	Time (s)	DC (%)	HD (px)	Time (s)
MD[9]	84.5	3.10	**1.5**	72.5	3.29	**1.0**
DL[12]	90.1	2.32	13.2	**93.0**	**2.09**	6.5
Ours	**90.2**	**2.31**	1.9	92.9	2.10	1.1

they aslo restrict the pose parameters to a certain domain.

6 Conclusion

We introduced a new transformation distance for shape registration in object segmentation, namely circulant dissimilarity. Since the object shape is approximately represented by a sparse linear combination, the transformation invariance (like translation or rotation) can be achieved by cyclic shift. Due to the circulant structure, shape transformation is represented as circular convolution, which could be implemented by using FFT. This circulant dissimilarity based shape registration and kernel density estimaton based shape priors are introduced in an energetic form to regularize the target shape in variational segmentation. In comparison with iterative optimization on explicit pose parameters, the proposed model could provide more reliable pose information and achieve satisfactory results even under large pose variability of the objects of interest.

Disclosures

The authors declared that they have no conflicts of interest to this work.

References

1 J. Yang, B. Price, S. Cohen, *et al.*, "Patchcut: Data-driven object segmentation via local shape transfer," in *Proc. CVPR*, (2015).

2 J. Wang, S. Yeung, and K. LukChan, "Matching-constrained active contours with affine-invariant shape prior," *Computer Vision and Image Understanding* **132**, 39–55 (2015).

3 E. Rodola, S. R. Bulo, and D. Cremers, "Robust region detection via consensus segmentation of deformable shapes," *Computer Graphics Forum* **33**, 97–106 (2014).

4 G. Hamarneh and X. Li, "Watershed segmentation using prior shape and appearance knowledge," *Image and Vision Computing* **27**, 59–68 (2009).

5 T. Cootes, C. Taylor, D. Cooper, *et al.*, "Active shape modelstheir training and application," *Computer Vision Image Understanding* **61**(1), 38–59 (1995).

6 S. Osher and R. Fedkiw, "Level set methods and dynamic implicit surfaces," in *New York: Springer-Verlag*, (2002).

7 M. Leventon, W. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. CVPR*, (2000).

8 D. Cremers, S. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for shape priors in level set segmentation," *Int. J. Computer Vision* **69**, 335 –351 (2006).

9 M. Rousson and D. Cremers, "Efficient kernel density estimation of shape and intensity priors for level set segmentation," in *Proc. MICCAI*, **3750**, 757–764 (2005).

10 F. Chen, H. Yu, and R. Hu, "Shape sparse representation for joint object classification and segmentation," *IEEE Trans. Image Processing* **22**(3), 992–1004 (2013).

11 V. Prisacariu and I. Reid, "Nonlinear shape manifolds as shape priors in level set segmentation and tracking," in *Proc. CVPR*, (2011).

12 F. Chen, H. Yu, R. Hu, *et al.*, "Deep learning shape priors for object segmentation," in *Proc. CVPR*, (2013).

13 M. Rousson and N. Paragios, "Shape priors for level set representations," in *Proc. ECCV*, (2002).

14 A. Tsai, A. Yezzi, W. Wells, *et al.*, "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE Trans. Medical imaging* **22**(2), 137–154 (2003).

15 T. Chan and W. Zhu, "Level set based shape prior segmentation," in *Proc. CVPR*, (2005).

16 S. Zhang, Y. Zhan, M. Dewan, *et al.*, "Sparse shape composition: A new framework for shape prior modeling," in *Proc. CVPR*, (2011).

17 A. Wagner, J. Wright, A. Ganesh, *et al.*, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach.Intell.* **34**(2), 372–386 (2012).

18 D. Cremers, F. Schmidt, and F. Barthel, "Shape priors in variational image segmentation: convexity, lipschitz continuity and globally optimal solutions," in *Proc. CVPR*, (2008).

19 J. Henriques, R. Caseiro, P. Martins, *et al.*, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. ECCV*, (2012).

20 R. M. Gray, "Toeplitz and circulant matrices: a review," in *Now Publishers*, (2006).

21 X. Zhou, X. Huang, J. S. Duncan, *et al.*, "Active contours with group similarity," in *Proc. CVPR*, 2969–2976 (2013).

List of Figures

2   Rotation invariance by cyclic shift in polar coordinates. A test shape (a) is converted from Cartesian coordinates to polar coordinates (b). By horizontal cyclic shift 30-degree along θ axes (c), a rotated shape (d) is obtained by using conversion from polar coordinates to Cartesian coordinates.

3   The shape registration (middle) is proposed to link low-level image data (left) with high-level shape prior (right) for object segmentation.

4   Typical segmentation comparison on tracking a walking person. The frames 5, 9, 23, 28, and 39 are shown in the first row. Segmentation results by iterative optimization on explicit pose parameters[8] are given in the middle row. Our segmentation results are presented in the last row.

5   Robust initialization. (a) The mean shape of all training shapes and its boundary (0.5-level set). (b) The different locations of initial curves were randomly chosen within a local range (yellow box). Two typical initial localtion were shown in (c) and (d). Noted that the initial locations were a certain distance from the walking person.

6   Selected sample shapes from the hand-posture dataset.

7   Typical comparison on hand-posture sequence 1. The original frames 3, 17, 22, 46, and 56 are shown in the first row. Segmentation results by iterative optimization on explicit pose parameters[9] are given in the second row. Segmentation results by exhaustive pose parameters search[12] are given in the third row. Our segmentation results are presented in the last row.

List of Tables