Non-Rigid Image Registration with MRF

Junwei Jason Zhang
Department of Computer Science
State University of New York at Stony Brook
Stony Brook, NY 11790
junweizhang23@gmail.com

Jian Jiang
Department of Computer Science
State University of New York at Stony Brook
Stony Brook, NY 11790
jianjiang@cs.stonybrook.edu

Abstract

The motivation of this project comes from the registration of 3D faces. Using Discrete Ricci Flow and conformal mapping, we can map a 3D face onto a 2D image. Registration work is done on this 2D image. Our project focuses on the second step. We use MRF energy to represent the matching problem and minimize the energy to achieve our goal. This work has been specified in [3].

1 Introduction

Face registration is always widely used in many areas. In medical field, cosmetology demands accurate registration between face models of the same person to see the change as time. And this market worths billions of dollars in United States.
According Prof. Xianfeng David Gu's previous work, a face 3D model could be mapped onto a 2D image using conformal mapping. Based on this work, we focus on the image registration using Markov Random Field(MRF). We define a MRF energy function and try to minimize it to get the deformation, which transforms an image to another. As for the optimisation method, we use both Iterative Conditional Modes(ICM) and TRW-S, and compare the results of them.
We will introduce the algorithm in Section 2, and present the result in Section 3. Future work will be found in Section 4.

2 MRF Energy Function

2.1 Define Energy Functioin

Since we use MRF to register two images, we need to define the MRF energy function.
Let $L = \{1...K\}$ be a set of labels. Let $G = (V, E)$ be a graph with $E \subseteq V \times V$. Each graph node is assigned with a label $x_s \in L$ and a labeling or configuration is defined as $\mathbf{x} = \{x_s | s \in V\}$. So the energy of a configuration $\mathbf{x}$ is defined as

$$E(\mathbf{x}|\theta) = \sum_{s \in V} \theta_s(x_s) + \sum_{st \in E} \theta_{st}(x_s, x_t) \tag{1}$$

where $\{\theta_s(i) \in \mathbb{R} | i \in L, s \in V\}$ is unary potentials and $\theta_s(x_s)$ is referred as the unary term. $\{\theta_{st}(i,j) \in \mathbb{R} | i, j \in L, st \in E\}$ is pairwise potentials and $\theta_{st}(x_s, x_t)$ is the pairwise term. The probablity distribution defined as

$$p(\mathbf{x}|\theta) \propto \exp(-E(\mathbf{x}|\theta)) \tag{2}$$

is a Gibbs distribution corresponding to a certain Markov Random Field(MRF). Thus the problem of maximize a posteriori(MAP) configuration corresponds to the energy minimization $\min_{\mathbf{x}} E(\mathbf{x}|\theta)$.

2.1.1 Deformation Definition

We first define the deformation between two images $I^1 : T^1 \to [0,1]^3$ and $I^2 : T^2 \to [0,1]^3$ where T^1, T^2 are sets of pixels. We define the configuation $\mathbf{x}$ as components

$$x_s = (x_s^1, x_s^2),\ s \in T^1 \tag{3}$$

representing a 2D displacement field over T^1. Coordinates x_s^1 and x_s^2 denote x and y displacements of pixel s, respectively, taking value from $L = \{K_{min}, ..., K_{max}\}$. So our deformation for some pixel I_s is defined as

$$D_{\mathbf{x}}(I_s) = I_{s+x_s} \tag{4}$$

For each pixel $s \in T$, we have a specific deformation. And this is we called *Non-Rigid Deformation*. Thus, we define the unary term of energy function as $\theta_s(x_s) = \frac{(I_s^1 - I_{s+x_s}^2)^2}{2\sigma_I^2}$, and the pairwise term is defined as $\theta_{st}(x_s, x_t) = \frac{||x_s - x_t||^2}{2\sigma_x^2},\ (s,t) \in \mathbb{E}$ where $\mathbb{E}$ is the set of all horizontally and vertically neighbouring pairs of pixels(See Fig. 1).

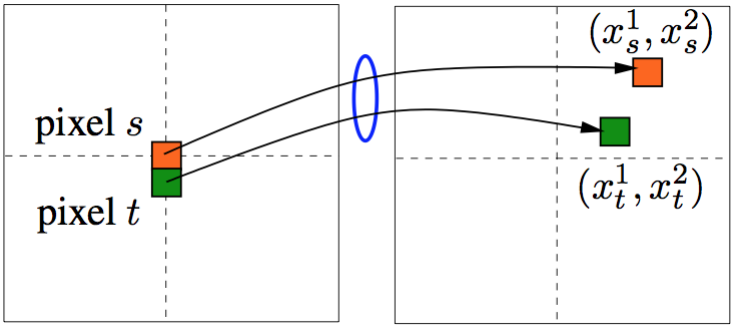

Figure 1: Deformation as pairwise MRF

2.1.2 Decomposed Model

We propose a model of *Decomposed Model* [2], representing x and y displacements by two interacting fields, referred as *layers*. Layer Φ^1 denotes x displacement, and Φ_2 denotes y displacement. Thus our configuration is defined as $\mathbf{x} = \{x_{s^i} | s^i \in \Phi^i,\ i = 1,2\}$. Then the unary term is

$$\theta_s x_s = \frac{\left(I_s^1 - I_{s+(x_{s^1}, x_{s^2})}\right)^2}{2\sigma_I^2} \tag{5}$$

And the pairwise term is

$$\theta_{st}\left(x_s, x_t\right) = \frac{\left(x_s - x_t\right)^2}{2\sigma_x^2},\ (s,t) \in \mathbb{E} \tag{6}$$

See Fig. 2.

2.2 Improvements

2.2.1 Block Model

If we minimize the energy over all pixels in the image, it will consume a lot of time. So we split the image into blocks to accelerate the compuation. Here we choose block size as 4 pixels. To avoid overfitting, we require the deformation field to be locally affine. As we consider discrete models, we want the deformation field to be described locally by translations. We propose to aggregate pixels into blocks and allow each block to have displacements with a pixel precision. We do not penalize relative displacements of ± 1 pixel in vertical and horizontal directions(see Fig. 3), and completely forbid larger displacements. So we define the unary term as

$$\theta_s(x_s) = \frac{1}{2\sigma_I^2} d(I_s^1, I_{s+(x_{s^1}, x_{s^2})}) \tag{7}$$

where $d(\cdot, \cdot)$ denots for the distance function of two blocks specified in the following. As for the pairwise term, it's defined as

$$\theta_{st}(x_s, x_t) = \begin{cases} 0 & x_s = x_t \\ C_r & |x_s - x_t| \leqslant 1 \\ \infty & |x_s - x_t| > 1 \end{cases} \tag{8}$$

where $0 < C_r \ll \frac{1}{\sigma_I^2}$ and we take $C_r = 0.001$ in implementation.

2.2.2 Distance of Blocks

3 Minimize Energy Function

3.1 ICM

Iterative Conditional Modes(ICM)[1] is firstly used to minimize our MRF energy function by us because it can be implemented simply. We will introduce its pipeline and then show the result of this optimization method.

3.1.1 ICM Pipeline

As for ICM, it traverses the whole label space for each graph nodes. First, we initialize the graph nodes with random labels, then the iteration begins. In each iteration, for every graph node, we

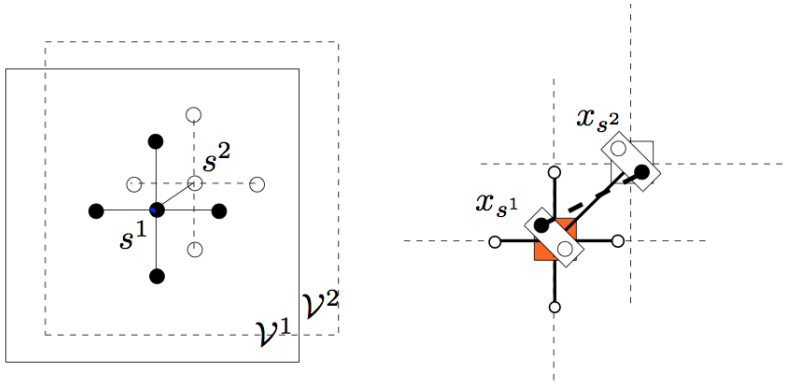

Figure 2: Decomposed Model

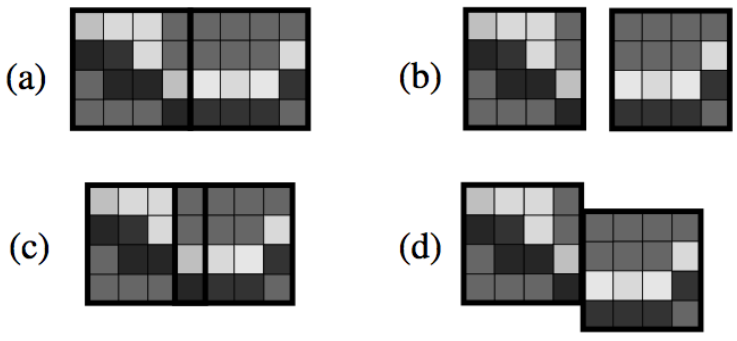

Figure 3: Block Model: (a) two neighboring blocks (b)-(d) examples which are not penalized

try every possible label for it, checking if the energy is smaller than previous configuration. If yes, we update the node with this label. Because we update the labels for nodes only when the energy under current configuration is smaller than previous configuration, it's guaranteed to converge. The pipeline is shown in Algorithm 1.

Algorithm 1 Pipeline for ICM

```
Initialize each node with a random label
while (Iteration Not Ends) do
    for Node i in all graph Nodes do
        for Possible Labels for node i do
            if (Energy is smaller than previous configuration) then
                Label for node i := label j
            end if
        end for
    end for
    Store current configuration
end while
```

3.1.2 Define the distance of colors

- For the color comparison we use

$$F_\lambda(c_1, c_2) = \lambda^2 \times \frac{\langle c_1 - c_2, c_2 \rangle}{\|c_2\|^2} + (c_1 - \langle c_1, c_2 \rangle \times \frac{c_2}{\|c_2\|}^2)^2$$

- We represent color space as $[0, 1]^3$ and compute color similarity using F_λ with $\lambda = 0.1$. For $d(x, y)$, we use

$$\frac{1}{N} \sum_{k=1}^{N} F_\lambda(x^i, y^i)$$

where x^i, y^i are colors of pixels in correspondence. We set $\sigma_I = 1$ and $c_r = 0.001$.

3.1.3 Speed up computation

- Using **rgb2ind** method to convert RGB space to 32 colors.
- Using mix-compile to maintain a Java Hash Table, contains $(c_1, c_2, d(c_1, c_2))$. In each computing, programming searches Hash Table first.
  If the table has the tuple(c_1, c_2), the Hash Table returns the $d(c_1, c_2)$.
  If not, the programming calculates the distance $d(c_1, c_2)$, depending on the definition and saves $(c_1, c_2, d(c_1, c_2))$.

3.1.4 Result

We use the human face emotion change to show our result(Fig. 4). We only show the blocks on both

Figure 4: Result of ICM. Left: the template face. Right: the target face

template face and target face on the boundary because it will cover the whole face if all blocks are shown. We can see from the result that the deformation of human faces is modeled very well and accurate under the circumstance where the emotion doesn't change too much. Because we restrict that neighboring blocks can't move apart too much, we can model small deformation very well. And since ICM searches for the whold label space and only updates labels when MRF energy is smaller, it's guaranteed to converge.
However, ICM has some shortages. First, it runs very slow. In our experiment[1], it takes about 30 minutes to compute the deformation for the image shown in Fig. 4, which has 335 blocks. In the future, it will much faster if we use parallel computing because we find that only a single core of CPU is utilized during the computation. Second, if may fall into local minimum.

3.2 TRW-S

3.2.1 TRW-S

T he tree-reweighted message passing (TRW) techniques, which approach the solution to MRF-LP via a dual problem defined by a convex combination of trees.

L ater, a sequential message passing scheme (known as TRW-S) was proposed. It updates messages in a sequential order instead of a parallel order used in TRW-E and TRW-T, which makes the lower bound will not decrease in TRW-S.

R egarding the convergence, TRW-S will attain a point that satisfies a condition referred to as weak tree agreement and the lower bound will not change any more since then.

R egarding the optimality, TRW-S cannot guarantee the global maximum of the lower bound in general.

3.2.2 Result

4 Summary and Future Work

References

[1] Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):pp. 259–302, 1986.

[2] I Kovtun. *Image segmentation based on sufficient conditions of optimality in NP-complete classes of structural labelling problem*. PhD thesis, PhD thesis, IRTC ITS National Academy of Sciences, Ukraine, 2004.(In Ukranian), 2004.

[3] Alexander Shekhovtsov, Ivan Kovtun, and Václav Hlaváč. Efficient mrf deformation model for non-rigid image matching. *Computer Vision and Image Understanding*, 112(1):91–99, 2008.

[1] Intel Core i7 2.7GHz, 8GB DDR3 RAM