Neurocomputing 120 (2013) 397–403

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Automatic cartoon matching in computer-assisted animation production

Zhijun Song [a,b], Jun Yu [c,*], Changle Zhou [a,b], Meng Wang [d]

[a] Cognitive Science Department, Xiamen University, Xiamen, China
[b] Fujian Key Laboratory of the Brain-like Intelligent Systems, Xiamen University, Xiamen, China
[c] Computer Science Department, Xiamen University, Xiamen 361005, China
[d] School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China

ARTICLE INFO

Article history:
Received 4 December 2011
Received in revised form
22 June 2012
Accepted 24 August 2012
Available online 26 March 2013

Keywords:
Matching
Maximum a posteriori
Maximum likelihood
Many-to-many correspondence

ABSTRACT

Traditional cartoon animation painting has always been a tedious job. In order to improve the efficiency of the process, the development of an automatic cartoon generation system including automatic inbetweening and coloring is required. Automatic matching of cartoon characters in key frames is the prerequisite for the system. This paper provides a novel matching algorithm with iterative maximum a posteriori (MAP) estimation and the maximum likelihood (ML) estimation. Specifically, this algorithm formulate cartoon matching as a many-to-many labeling problem. To refine the results of matching, an optimization approach is adopted to alternatively conduct the MAP estimation and the ML estimation. Besides, we construct the correspondence by using the local shape descriptor, and the rotation and scale invariance in matching can be achieved. The experimental results on real-world datasets demonstrate the effectiveness of the proposed methods for automatic cartoon matching.

1. Introduction

In conventional cartoon production, users conduct tedious procedures of inbetweening and coloring to generate inbetween frames and fill colors in them based on the given key frames. To leverage computing power for efficient animation production, automatic cartoon synthesis, inbetweening and coloring functions have been added into the computer-assisted 2D animation systems.

For cartoon synthesis, Yu et al. [43] provided a novel cartoon reusing framework to synthesize cartoons based on manifold learning [37–41,46]. This framework adopts multiple features of color, edge and shape to describe cartoon characters, and synthesizes novel cartoon clips from the existing cartoon datasets. However, this method cannot generate new frames in the synthesized cartoon clips. Unlike the idea of cartoon synthesis, Fekete et al. [9] provided a correspondence construction based inbetweening approaches which can generate new cartoon frames through interpolations. In addition, Kort [16] proposed an interactive model for automatic inbetweening with a possible solution for "undetectable" parts noticed in [9]. However, this approach is proposed for relatively simple animation styles such as cut-out animation. Chen et al. [8] offered a similar approach which applied heuristic method on feature points. Since the matching method cannot achieve global optimization, the effectiveness of correspondence construction is limited. Seah et al. [20,31] proposed the MFBA algorithm involving matching key frames and automatic inbetweening for raster images. However, this method does not perform well to keep the shape and smoothness of outlines for other types of images like cartoon characters. To solve the problem of matching complicated cartoon characters, Yu et al. [42] proposed a novel correspondence construction framework based on semi-supervised graph-based learning [10,11,33–35,44,45]. Specifically, the new framework constructs local patches for each point on an object and aligns these patches in a new feature space, in which correspondences between objects can be detected by the subsequent clustering. Besides techniques proposed in cartoon animation, related works on correspondence construction can be found in shape matching [1,23], shape registration [47], shape alignment and shape retrieval [22].

The first stage of matching is to represent cartoon characters through appropriate descriptors. A large amount of descriptors [1,19,21,24–26] have been proposed to describe shapes. However, in animation production, characters in key frames usually have complicated structures with open shapes. Thus, the shape descriptor-"shape context" [1,23] is suitable for our application because it uses related distribution to model complex character. For matching methods, the Hungarian algorithm [27] is applied to shape context matching [1]. However, it can only obtain one-to-one correspondence, so how to equalize the number of sampling points becomes a problem. Singular value decomposition (SVD) based methods [28,30] share the same problem. Early spectral methods on correspondence construction, such as [32,36], directly apply eigen-decomposition on an adjacent matrix. Because of high dimensionality and data redundancy, these methods are not

* Corresponding author. Tel.: +86 592 2580033; fax: +86 592 2580035.
E-mail address: yujun@xmu.edu.cn (J. Yu).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.08.051

Table 1
Important notations used in this paper.

Notation	Description
$p_s \in P$	The source character for correspondence construction. $p_s \in P$ represents each feature point in this character.
$q_t \in Q$	The target character for correspondence construction. $q_t \in Q$ represents each label.
f_p	The labels configuration for each point in source character.
$\phi(i)$	The feature representation for each label.
$\Phi = \{\phi(p_s)\}$	The union of the predicted corresponding labels for the point in source character.
$I(p_s)$	The 60-D shape descriptor for the feature point p_s.
$D(\cdot)$	The data penalty term in energy function.
$V_{p_i,p_j}(\cdot)$	The smoothness penalty function in energy function.

robust and accurate. Moreover, similar to Hungarian algorithm and the SVD based methods, they can only obtain one-to-one correspondence. To achieve many-to-many correspondence, different clustering techniques [5–7] are proposed. However, these methods' performance is seriously affected by the size of clusters which cannot be controlled by users in most cases.

Recently, a group of novel techniques have been developed based on graph cuts [3,15]. The basic technique of using graph cuts is to construct a specialized graph for the energy function to be minimized such that the minimum cut on the graph also minimizes the energy. These methods have been successfully used for a wide variety of applications including image segmentation [4,18,29], image restoration [2,3], image synthesis [17], multicamera scene reconstruction [14] and medical imaging [12,13]. To take image segmentation as an example, the purpose of the segmentation task is to assign each label to a group of pixels in the image. Thus, the labels and the pixels can form a graph in which the graph cuts can be used to obtain the optimal one-to-many assignment. Accordingly, the correspondence construction can be assumed as a labeling problem. Given two characters S and P, the many-to-many correspondence can be built by assigning the feature points $s_i \in S$ to the feature points $p_i \in P$, and vice versa. Specifically, to estimate the feature point assignment, we adopt an iterative optimization scheme to alternatively carry out the maximum a posteriori (MAP) estimation and the maximum likelihood (ML) estimation. The MAP estimation problem is modeled with Markov random fields (MRFs) and the graph cuts algorithm is used to find the solution to the MAP estimation. The ML estimation is achieved by computing the means of feature points in a Gaussian model. In summary, we proposed a novel framework of many-to-many cartoon correspondence construction. Correspondence construction is formulated as a label assignment problem through the maximum a posteriori (MAP) estimation, in which the graph cuts method is adopted to obtain the many-to-many results. By integrating the maximum likelihood (ML) estimation with MAP procedure, the correspondence construction result can be iteratively refined into an optimal one.

2. Automatic cartoon matching by label assignment

The input of our method is a pair of key frames described by the global shape context descriptor [1]. Two shapes P and Q can be defined by feature points $\{p_1, p_2, \ldots, p_n\}$ and $\{q_1, q_2, \ldots, q_m\}$, respectively. The shape descriptor of p_s is defined as a histogram h_s of the remaining $n-1$ points of the shape P under a coordinate system: $h_s(c) = \#\{p_j: j \neq s,\ p_j - p_s \in bin(c)\}$, in which the bins uniformly divide the log-polar space. However, it cannot be directly applied in our application because of the global shape context's invalidation in rotation invariance. The global shape context descriptor can be modified by using both the related distribution and orientation in feature extraction [1], and it is called oriented shape context [1].

Given a source character P, each point $p_s \in P$ is defined by the 60-dimensional oriented shape context feature $I(p_s) = [I_1(p_s), \ldots, I_{60}(p_s)]$. In carton matching, the target character $Q = [q_1, \ldots, q_m]$ should be provided. If we formulate it as a labeling problem, the m points in character Q can be assumed as the labels, and the objective is then to find a label configuration $f = \{f_p | p\}$. The label f_p denotes which point in Q the points in P is grouped into. For a specified label q_i, there are k points in P grouped into. Then, a 60-dimensional vector can be constructed as

$$\phi(i) = I(q_i) + \sum_{j=1}^{k} p_j, \tag{1}$$

which is used to describe the properties of label q_i. The 60 components have the similar meanings to those of the corresponding 60 components of $I(p_s)$. Let $\Phi = \{\phi(p_s)\}$ be union of the predicted corresponding labels for $p_s \in P$. If P and Φ are known, the matching is to find an optimal correspondence configuration f', which maximizes the posterior possibility of the label configuration

$$f' = \underset{f}{argmax}\ Pr(f | \Phi, P), \tag{2}$$

where f' can be obtained by either a learning process or an initialized estimation. The solution of the label configuration is introduced in Section 3. For convenience, Table 1 lists the important notations used in the rest of the paper.

3. The framework of cartoon matching

Fig. 1 shows the framework of automatic cartoon matching. In step 1, a pair of input key frames is generated. In step 2, these two frames are sampled as two points sets $P = [p_1, \ldots, p_n]$ and $Q = [q_1, \ldots, q_m]$, and these points are modeled by the oriented shape context in step 3. Afterward, in step 4, the points $p_s \in P$ are denoted as the feature points and the points $q_t \in Q$ are denoted as the labels. And the MAP estimation is conducted to build a graph representation. In step 5 and step 6, the approaches of graph cuts [3] and ML estimation are iteratively conducted to obtain the optimal label configurations. Finally, in step 7, the stroke correspondence is built based on the point correspondence.

3.1. MAP estimation of label configurations

Based on the character P and the potential label features Φ, the labels f can be inferred by the Bayesian law, i.e., $\Pr\left(f \middle| \Phi, P\right)$ can be obtained by

$$\begin{aligned} \Pr(f | \Phi, P) &= \frac{\Pr(\Phi, P | f)\Pr(f)}{\Pr(\Phi, P)} \\ &\propto \Pr(\Phi, P | f)\Pr(f), \end{aligned} \tag{3}$$

which is a MAP estimation problem and can be modeled by MRFs. Based on the independent identical data distribution, $\Pr(\Phi, P | f)$ can