

智能装配视觉系统新的三维重建方法^{*}

熊银根 张光昭

Paul S Wu

(中山大学无线电电子学系, 广州 510275) (香港城市大学)

摘 要 提出一种新的三维重建方法, 对所获得的立体象对进行适当的变换, 把空域中的立体匹配问题转换到波数域中来解决; 采用基于局部相的立体匹配方法来求解立体匹配这个难题, 并用 Hopfield 神经网络来实现立体匹配过程, 以便实现并行处理; 构造了一个三维重建神经网络来实现三维重建。

关键词 智能装配, 计算机视觉, 三维重建, 立体匹配, 神经网络

分类号 TP 205

与二维物体识别不同, 三维物体识别需要被识别物体的三维信息。为了获得这些三维信息, 通常是利用立体象对, 对其进行立体匹配后再进行三维重建就能获得物体的三维外形, 从而进行三维物体识别, 然而, 在此过程中, 立体匹配是立体视觉中一个最难解决的问题, 也是三维信息提取与重建中最关键的问题, 有关这方面的研究已做过大量的工作^[1-4], 也取得了许多成果。由于问题本身的复杂性和实际情况要求的不同, 仍存在着很多难以解决的问题。为了满足智能装配对视觉系统进行三维物体识别的要求, 本研究提出了一种新的立体匹配方法, 该方法是把立体匹配问题转换到波数域中进行求解, 在波数域中求出图象的局部相后, 根据局部相的性质, 建立相应的匹配函数。为了提高匹配过程的处理速度, 易于实现并行操作, 在本研究中采用 Hopfield 神经网络来实现立体匹配过程^[5], 为此, 建立了满足一定约束条件的能量函数, 对其进行极小化就可以得到图象的最佳立体匹配。在解决了立体匹配问题的基础上, 本文还提出了利用神经网络来进行物体三维外形重建。

1 三维重建的方法

图 1 是三维重建方法的总体结构。其大体过程是获取立体象对, 对立体象对进行预处理, 对立体象对的左右图象进行适当的变换, 求出其局部相; 利用 BP 神经网络, 估计立体象对的两幅图象之间的对应核线, 利用 Hopfield 神经网络进行基于局部相的立体匹配过程; 最后是利用 BP 神经网络进行三维物体外形重建。

1.1 图象局部相的计算

在获取待识别零件的立体象对后, 对其进行预处理, 采用阈值函数把零件与背景分开。

^{*} 广东省博士后基金资助项目

收稿日期: 1996-07-11 熊银根, 男, 34岁, 博士后

在立体匹配之前,必须提取图象的特征,这个特征通常是图象的边缘,这种信息含有一定的模糊性.在波数域中我们不需要特征提取过程,立体匹配过程中很多在空域中遇到的问题,在波数域中将可以避免. Fourier 变换能把空域问题转换成波数域问题,因此,可以在波数域中来研究输入信号的有关性质.

对于一个连续的一维信号 $f(x)$,其 Fourier 变换为 $F(u)$,它的位移为 $f_s(x) = f(x-s)$. 因此,与空域中的位移相对应的波数域中的相位移为

$$F_s(u) = \int_{-\infty}^{\infty} f_s(x) \exp(-j2\pi ux) dx = \exp(-j2\pi us) F(u) \quad (1)$$

式中, u 为波数, s 为空域中的位移.

为了表示波数域中的相移与空域中的位移的关系,定义

$$b(x) = \exp(j2\pi ux) \times f(x) = \int_{-\infty}^{\infty} f(t) \exp[j2\pi u(x-t)] dt = \exp(j2\pi ux) F(u) = |b(x)| \exp[\arg[b(x)]] \quad (2)$$

其中

$$\arg[b(x)] = 2\pi ux + \arg[F(u)]$$

$$\text{同样 } b_s(x) = \exp(j2\pi ux) \times f_s(x) = \exp[j2\pi u(x-s)] F(u) = |b_s(x)| \exp[\arg[b_s(x)]]$$

其中

$$\arg[b_s(x)] = 2\pi ux + \arg[F(u)]$$

因此

$$\arg[b(x)] - \arg[b_s(x)] = 2\pi us \quad (3)$$

从式 (3) 中可以看出,空域中的位移 s 能在波数域中用相移探测到.

然而, Fourier 变换是整个信号的全局处理,它只能反映唯一的全局位移,位移函数沿 x 轴方向是常数,并且相的变化是严格线性的.在立体匹配问题中,视差在整个图象上并不一定是常数,这里需要图象的局部信息,因此, Fourier 变换对提取 $f(x)$ 的局部频谱信息没有作用.

加窗 Fourier 变换能用来分析信号的局部信息,图象的视差能用局部相移来表示.我们定义一个矩形窗函数,其参数为 M

$$W_M(x) = \begin{cases} 1 & \text{当 } |x| \leq M/2 \\ 0 & \text{其它} \end{cases}$$

这时,加窗 Fourier 核为 $h(x) = W_M(x) \exp(j2\pi ux)$.

因此,式 (2) 中的卷积变成

$$g(x) = h(x) \times f(x) = \int_{-\infty}^{\infty} W_M(t) f(x-t) \exp(j2\pi ut) dt = \int_{-M/2}^{+M/2} f(x-t) \exp(j2\pi ut) dt \stackrel{\text{def}}{=} R(x) + jI(x) = - \int_{-M/2}^{+M/2} f(x+s) \exp(-j2\pi us) ds$$

$$\text{其中: } R(x) \text{ 为 } g(x) \text{ 的实部 } R(x) = \int_{-M/2}^{+M/2} f(x-t) \cos(2\pi ut) dt \quad (4)$$

$$I(x) \text{ 为 } g(x) \text{ 的虚部 } I(x) = \int_{-M/2}^{+M/2} f(x-t) \sin(2\pi ut) dt \quad (5)$$



图 1 三维重建方法的总体过程

Fig. 1 The structure of 3D reconstruction process

因此, $g(x)$ 的幅值为 $|g(x)| = [R^2(x) + I^2(x)]^{1/2}$.

加窗 Fourier 相定义为

$$h(x) = \arg[g(x)] = \arctan2[I(x), R(x)] \quad (6)$$

式中, $\arctan2(y, x)$ 与 $\arctan(y/x)$ 相同, 但值域为 $(-\pi, \pi]$.

在利用加窗 Fourier 变换求解局部相之前, 一个很重要的工作是去除噪声, 通过采用高斯滤波

$$n(x) = (1/\sqrt{2\pi}) \exp[-(x^2/2\sigma^2)]$$

信号变成

$$g(x) = h(x) \times [n(x) \times f(x)] = [h(x) \times n(x)] \times f(x)$$

Gabor 滤波也能用来获取信号的局部信息, 它能用相对较大的视差来对应高频信号^[7].

加窗 Fourier 变换能扩展到处理二维信号, 把 Fourier 核表示成二维函数, 并且定义为二维窗口函数. Fourier 核与窗口函数进行卷积, 能得到二维加窗 Fourier 核, 通过信号与二维加窗 Fourier 核进行卷积能获得二维加窗 Fourier 相.

在空域中, 我们使用图象的特征或纹理来进行匹配过程, 在波数域中, 立体匹配问题可以构造局部相的匹配, 这就是在立体象对的左右图象之间, 其对应点处的局部相是相等的, 即如果 x_l 为左图象上的点, 其对应右图象上的点为 x_r , 则有

$$h(x_l) = h(x_r) \quad (7)$$

这是构造立体匹配算法的基础.

1.2 基于局部相的匹配

在进行立体匹配过程之前, 必须估计立体象对的核线, 它们能为立体匹配过程提供很有用的约束, 投影在左图象核线上的点, 必定会投影到右图象相应点的核线上. 本研究采用一个 BP 神经网络来估计立体象对的左右图象之间的对应核线. 该神经网络有 4 层, 在输入层有 3 个神经元, 分别对应于左图象的 x 和 y 值, 以及右图象的 x 值. 具有两个隐含层, 每个隐含层具有 20 个神经元, 用来贮存左右图象之间的关系, 输出层具有一个神经元, 对应于右图象的 y 值.

传统的立体匹配方法一般是对图象的特征或纹理在空域中进行匹配, 与此不同, 这里, 我们在波数域中进行求解, 利用局部相的性质能找出左右图象之间的对应关系.

立体匹配是立体视觉中最难解决的问题, 而且需要很长时间才能完成. 然而, 在很多实际情况下, 视觉系统是在线操作, 要求准确而且尽可能快地识别有关物体, 因此, 要求立体匹配过程尽可能快.

大规模并行处理是神经网络最大的优点之一, 为了尽快完成立体匹配过程, 我们构造一个神经网络来实现立体匹配过程. 这里是采用 Hopfield 神经网络来组成这一结构的, 神经网络的层代表图象从上到下的核线, 在每一层上的行代表左图象核线上从左到右的像素点, 列代表右图象核线上从左到右的像素点, 每一个神经元代表一个可能的匹配.

为了采用 Hopfield 神经网络来求解立体匹配问题, 我们必须定义一个能量函数, 其极小值对应于问题的一个最好解, 在立体匹配过程中, 必须满足一定的约束, 能量函数就是使用这些约束构造出来的. 这些约束有: 唯一性、核线、顺序、连续性和几何约束, 为此, 首先定义一个匹配函数.

$$\Delta h_{x,i} = [h(X, I) - h(i, I)]^2 \quad (8)$$

其中: $\triangle h_{X,i}$ 为左图象上点 (X, l) 与右图象上点 (i, l) 处的相差; $h(X, l)$ 为左图象点 (X, l) 的相; $h(i, l)$ 为右图象点 (i, l) 的相; l 为第 l 条核线.

在理想情况下有: $\triangle h_{X,i} = 0$.

实际上, 由于系统误差, 上式不可能得到严格满足, 我们用极小化匹配函数 $\triangle h_{X,i}$ 来代替上式

$$\min[h(X, l) - h(i, l)]^2 \quad (9)$$

根据上述的有关约束, 能量函数可以表达为

$$\begin{aligned} E = & \frac{A}{2} \sum_l \sum_i \sum_X \sum_{Y \neq X} V_{l,X,i} V_{l,Y,i} + \frac{B}{2} \sum_l \sum_X \sum_i \frac{1}{N_{nor}} \sum_{j \neq i} [h(X, i) - h(j, l)]^2 V_{l,X,i} V_{l,X,j} \\ & - \frac{C}{2} \sum_l \sum_X \sum_i \frac{1}{N_{nor}} \sum_{Y < X} \sum_{j > i} V_{l,Y,i} V_{l,X,i} + \sum_{Y > X} \sum_{j < i} V_{l,Y,j} V_{l,X,i} - \\ & \frac{D}{2} \sum_l \sum_X \sum_i \frac{1}{N_{nor}} \left(\sum_{Y \in s} V_{l,X,i} V_{l-1,Y,j} + \sum_{Y \notin s} V_{l,X,i} V_{l-1,Y,j} \right) + \\ & \frac{E}{2} \sum_l \sum_X \sum_i d_{l,X,i} V_{l,X,i} + \sum_l \sum_X \sum_{i > X} 10000 V_{l,X,i} \quad d_{l,X,i} > X \end{aligned} \quad (10)$$

式中: l 为层号; X, Y 为左图象核线上第 X 个和第 Y 个像素; i, j 为右图象核线上第 i 个, 第 j 个像素; $d_{l,X,i}$ 为第 l 层上像素点 (X, l) , (i, l) 处的匹配函数值; s 为矩形区域; X 为阈值.

对上述能量函数进行极小化, 实现立体匹配.

1.3 物体三维外形重建

通过立体匹配, 我们获得了立体象对的左右图象之间的对应点, 为了获得三维物体外形, 必须从图象上的二维点计算出物体的三维点, 进行三维外形重建. 这里采用一个 BP 神经网络把匹配后的二维点转化成空间三维点. 该神经网络有 4 层, 在输入层有 4 个神经元, 它们分别对应左右图象上匹配后对应点的 x, y 坐标, 有两个隐含层, 每个隐含层有 20 个神经元, 在输出层具有 3 个神经元, 分别对应目标点的空间坐标 X, Y 和 Z 值.

该三维重建神经网络通过用样本进行训练, 可以建立系统中二维图象与三维物体之间的关系. 因此, 在使用这个神经网络之前, 很重要的一步是用样本对其训练. 隐含层的神经元是用来贮存通过训练所学习到的系统复杂关系的, 训练的精度直接影响到计算三维点坐标的结果.

为了实现并行处理, 这里可以采用多个神经网络同时进行从二维点到三维点的计算, 这样可以减少处理时间, 为了获得完整的物体外形, 可以对所得到的结果进行插值拟合.

2 实验及其应用研究

为了研究本文所提出的理论和方法的正确性和实用性, 这里用一些实例来进行检验. 第一个实例是用计算机产生的随机点图 (图 2 a, b), 其视差具有 4 个级别, 分别为 0, 8, 16, 24. 从立体镜中我们可以看到, 其立体形状为一个“婚礼蛋糕”, 图 2(c) 为用本文所提出的三维重建方法所得出的结果, 可以看到, 其三维形状得到了很好的恢复.

图 3(a, b) 是神经视觉系统所获取的一个立体象对, 在场景中的零件是一个齿轮, 我们要利用这个立体象对提取其三维信息, 重建其外形表面, 以便进行三维零件识别.

要进行三维外形表面重建, 首先必须进行立体匹配, 图象经过立体匹配后, 获得了左图图象上相对应的点. 利用这些匹配结果进行三维外形重建, 利用三维重建神经网络, 可以求

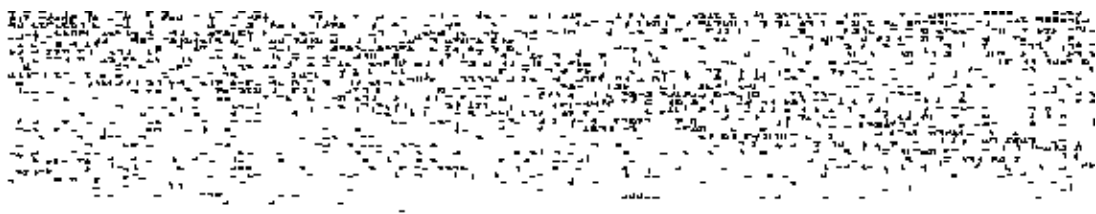


图 2 “婚礼蛋糕”随机点图 (a, b)和重建后形状 (c)

Fig. 2 Random-dot stereogram(a, b) and 3D reconstruction for wedding-cake(c)

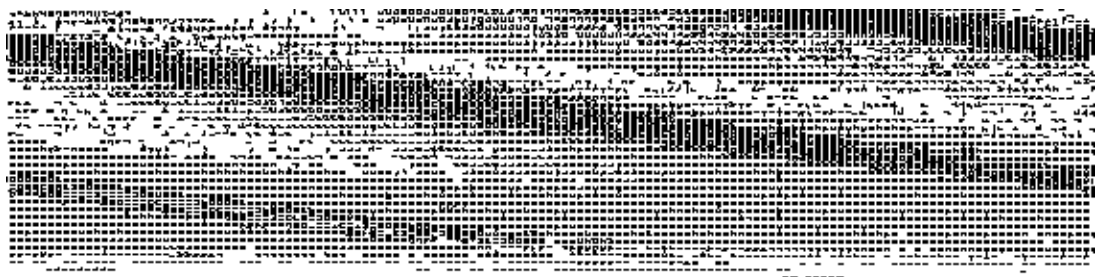


图 3 齿轮立体象对 (a, b)和重建三维外形 (c)

Fig. 3 Stereo image pair of gear (a, b) and 3D reconstruction(c)

出物体表面的三维坐标,图 3(c)为重建出来的三维外形表面。

从该结果中可以看出,齿轮的三维外形也得到了很好的重建。此外,本研究还把该方法进行了其它应用,都收到了较好的效果。

3 结 论

本文为了满足三维零件识别的需要,提出了一种新的三维重建方法,与传统方法不同,该方法把立体匹配问题转换到波数域中来进行求解,对立体象对进行适当地变换后求出每幅图象的局部相,根据局部相的性质,求出立体象对的两幅图象之间的对应关系。为了便于实现并行处理,采用 Hopfield 神经网络来实现立体匹配过程。采用 BP 神经网络构造成三维重建神经网络,把图象上的二维点转换成物体表面上的三维点,避免了传统方法中需要对视觉系统进行精确校准来确定象平面与目标空间的关系,以及复杂的迭代过程来计算物体的三维点。本文对所提出的方法进行了实验验证和应用研究,无论是对合成图象还是自然图象都收到了很好的效果,证明该方法是正确的,能满足智能装配系统中三维零件识别的要求。

参 考 文 献

- 1 Medioni G, Nevatia R. Segment-base stereo matching. *Computer Vision Graphics Image Processing*, 1985, 31: 2~ 18
- 2 Ito M, Ishii A. Three view stereo analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1986, PAMI-8(4): 524~ 532
- 3 Kanade T, Okutomi M. A stereo matching algorithm with an adaptive window: theory and experiment. *Proceedings of the 1991 IEEE International Conference on Robotics and Automation* Sacramento, California, 1991. 1088~ 1095
- 4 Steven D C, Medioni G. 3-D surface description from binocular stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14(10): 981~ 994
- 5 Yi-Tong Z, Chellappa R. *Artificial neural networks for computer vision*. Springer-Verlag, 1992. 6~ 34
- 6 Yalin X, Shafer A S. Variable window gabor filters and their use in focus and correspondence. *Proc Computer Vision Patt Recg*, 1994. 668~ 671
- 7 Sanger T. Stereo disparity computation using gabor filters. *Biol Cybern*, 1988, 59: 405~ 418
- 8 Yingen X, Qiao X, Dezong W. The measurement and three dimensional reconstruction for aircraft using image processing techniques. *Transactions of Nanjing University of Aeronautics & Astronautics*, 1993, 10(1): 216~ 219

New 3D Reconstruction Approach for Intelligent Assembly Vision System

Xiong Yingen^{} Zhang Guangzhao Paul S Wu*

Abstract A new 3D reconstruction approach is presented. First, the stereo image pair obtained by vision system is transformed, and then the local phase is obtained for stereo matching processes. The Hopfield neural network is established, so that the computation can be implemented efficiently in parallel. A 3D reconstruction neural network is constructed by using BP neural network. The 2D points on images can be converted into 3D points on objects by using 3D reconstruction neural network. At last, the examples for both synthetic and real image are shown in the experiment, and good results are obtained.

Keywords intelligent assembly, computer vision, 3D reconstruction, stereo matching neural network

^{*} Department of Radio and Electronics, Zhongshan University, Guangzhou 510275