1. **手语合成研究工作主要进展**

基于视频拼接的手语合成方法是采用视频录制设备采集手语词视频数据，构建手语词视频数据库并建立索引，针对所要表达的语句在手语数据库中搜索对应词汇，通过对手语词视频片段重组合成手语视频。现有的基于视频拼接的手语合成技术存在合成结果不够平滑，自动化程度低，不便于大数据量处理等问题。本项目，我们利用Kinect采集的手语视频原始语料库，结合图像处理和图形处理技术，研究更为自动、真实感更强的手语合成。从立项至今，具体完成工作如下。

**2.1基于视频拼接的手语合成**

**2.1.1手语过渡帧构建算法**

过渡帧就是从前一个姿势到后一个姿势的中间运动过程，而人体运动可以认为是由骨架控制的。为了得到更真实，更符合人体运动规律的过渡帧，用预处理得到的过渡骨架控制过渡帧变化。现有的图像插值技术并不能处理有大幅度动作变化的场景。我们将形变技术应用到过渡帧构建中，研究适用于手语合成的插值方法。在上年度基于图像形变的手语过渡帧插补方法的基础上，我们提出了一种改进的图像插补方法——基于形变辅助的图像插补算法。

**基于形变辅助的图像插补算法**

形变辅助的图像插补算法流程图如图 5.3所示。给定源图像和，插补目标是生成中间过渡帧（设定和）。形变辅助的插补算法可以总结为以下两步。首先，根据用户设定的动作路径（在手语合成中就是数据预处理生成的骨架过渡路径）生成形变序列。是由图像形变生成的前向形变序列，是由图像形变生成的后向形变序列。然后，采用一种基于块的图像融合算法融合前向形变序列和后向形变序列，得到满足源相似性和时间相关性的最终中间过渡帧。换言之，每一帧中间图像不仅要相似于相关的形变图像和，也要相似于时间相邻帧图像和。

将图像融合问题构建为一个优化问题，目标方程如下：

。

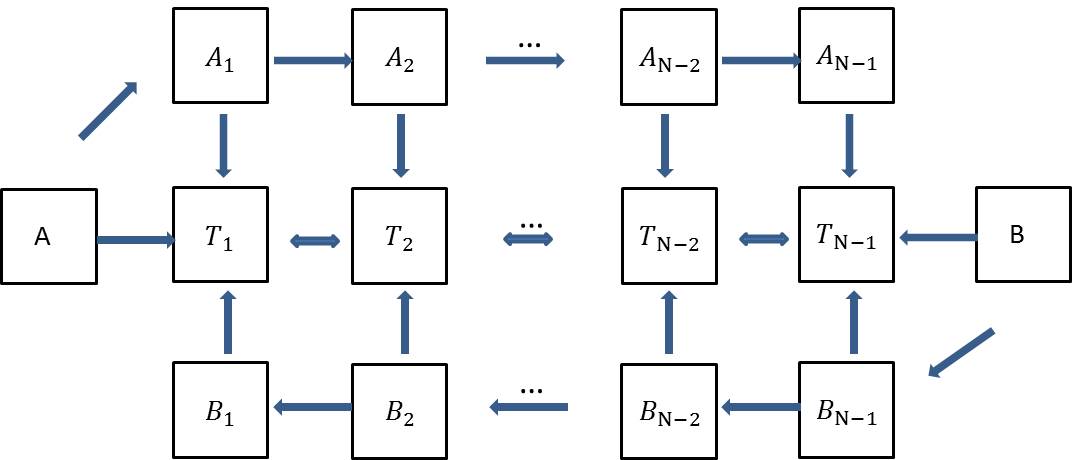
，，，分别是，，，的相似权值，目标方程(5.28)式包含源相似性和时间相关性。这里，用双向相似距离（bidirectional similarity，BDS）【1】来描述目标过渡图像与各参考图像之间的关系。BDS距离包含两方面度量：完整性度量和一致性度量。完整性度量要求源图像中的图像块包含在目标图像中，保证目标图像包含尽可能多的源图像信息；一致性度量衡量目标图像中的图像块与参考图像的相似性，避免产生源图像中没有的信息和不自然的图像结构。图像和的BDS距离定义为：

，

，

。

和分别是尺寸为在图像和中的图像块，是梯度算子。计算CIELAB颜色空间下图像块的平方和距离，计算范数距离，平衡梯度距离和颜色距离的权重。因为实际图像的梯度值是稀疏的，范数对梯度距离的度量，能够更好地保护图像的细节和纹理。完整性度量就是在中寻找与每一个图像块最相似的块，并计算所有对应图像块之间的距离之和。类似地，一致性度量就是在中寻找与每一个图像块最相似的块，并计算所有对应图像块之间的距离之和。



**图 5.3** 形变辅助的插补算法流程图

求解目标方程的过程可以概括为先单独求每个源图像的最佳相似图像再融合。单个源图像的最优匹配图像采用EM算法迭代求解。

最优匹配图像优化问题：

。

EM迭代求解目标图像的过程可以总结为：在E步，在源图像中检索出与目标图像中的图像块最匹配的图像块，集合所有重叠图像块的像素值；在M步，用重叠图像块的像素值混合出目标图像的像素。为了避免陷入局部最优，采用尺度变化的求解策略。也就是先从低分辨率的图像开始EM算法迭代，逐步改变图像尺寸，在新的图像尺寸下重新进行EM迭代，直至达到原始图像分辨率。

（a）块搜索

块搜索就是检索源图像中与目标图像中的图像块最相似的图像块，搜索空间有平移，旋转，尺度变换三个自由度。为了避免细微的亮度、颜色变化对搜索结果的影响，引入增益和偏差调整图像块【2】：

，

，

。

上式中，代表颜色通道，是图像块的标准差，是图像块的均值，与是增益与偏差的范围。

块搜索的核心在于计算图像块的相关性。给定源图像中的像素点，其在目标图像中的对应像素点为。最近邻场（nearest neighbor field，NNF）代表图像间像素点的位移。块匹配算法可以总结为以下过程：首先，用随机位移和先验信息初始化最近邻场；接着，通过一个迭代更新的过程更新最近邻场。这个迭代更新过程包括块位移向邻近像素传递以及随机搜索最优匹配。

（b）“投票混合”

利用上一步检索到的图像块得到最优目标图像：

，

上式中，是目标图像中的图像块，是块搜索得到的的最近邻图像块。由于平方距离可以转化为：

，

这里是变量的常数方程。因此可以用最优图像块混合出目标像素：

。

上式中，和是由参考图像中的像素混合得到的，位置的值分别为：

，

。

这里是源图像中距离目标图像块最近的图像块（是图像块左上角的坐标），选择图像块中处的像素。像素点的颜色和梯度分别是覆盖位置图像块的颜色和梯度平均值。

对于目标函数式，分别通过上述的重构方法求解出每一项，可以得到和，则通过最小化下式得到第帧的目标图像帧：

，

，

。

目标颜色值可以通过线性混合得到：

。

由于用范数衡量梯度距离，则当最小时至少有一项为零值。因此，选取中梯度幅度最大的作为梯度值：

。

最后将求解到的颜色值和梯度值用泊松重构算法计算出最终的过渡帧，可以将问题构造成以下形式：

，

上式中是泊松重构的目标图像，和分别是EM迭代块匹配得到的颜色和梯度。用欧拉—拉格朗日方程求解，方程转化为：

，

。

用傅里叶变换求解泊松重构问题，将空间频域定义为，对式两边做傅里叶变换得到：

。

，，和分别是，，和的傅里叶变换。因此，可以得到的傅里叶形式：

。

最终，可以通过快速求解泊松重构问题得到目标图像。在本算法中，各源图像的权重与图像帧的位置相关，假设总帧数为，则第帧的各项权重为：

。

结合前文所述的预处理和局部控制的形变算法，用形变辅助的图像插补处理中间过渡帧得到结果，结果如下图所示。

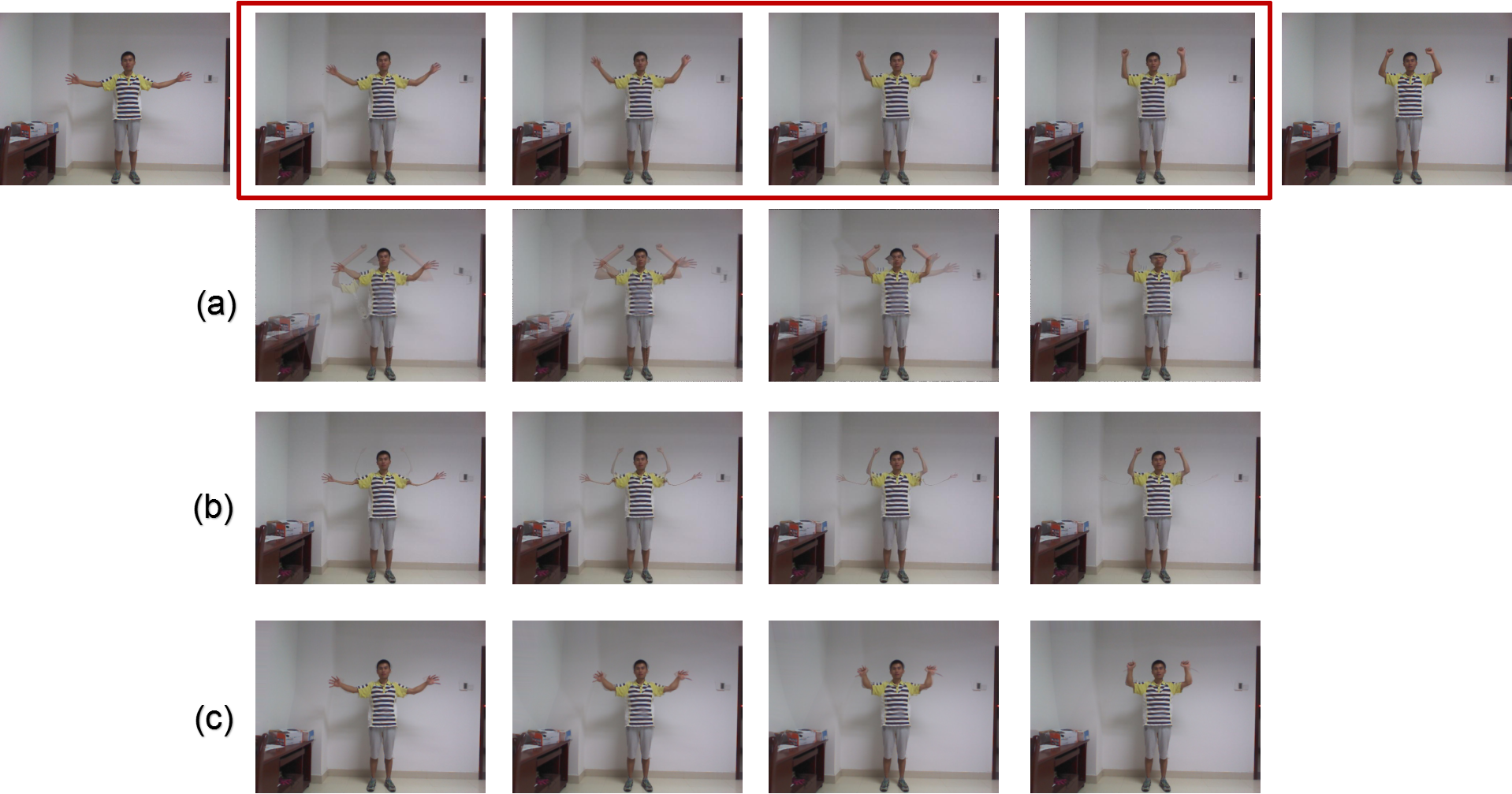


图 形变辅助的图像插补的手语过渡帧合成

[1] Simakov D, Caspi Y, Shechtman E, et al. Summarizing visual data using bidirectional similarity[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[2] HaCohen Y, Shechtman E, Goldman D B, et al. Non-rigid dense correspondence with applications for image enhancement[J]. ACM transactions on graphics (TOG), 2011, 30(4): 70

**2.2基于模型的手语合成**

基于模型的手语合成是以手语数据库的手语词数据为基础，对虚拟人进行3D建模，针对所要表达的语句在手语数据库中搜索对应词汇，根据手语动作参数对运动轨迹进行估计，进而控制虚拟人运动来合成手语动画。手语过渡词汇间过渡的平滑手语合成质量评价的关键特征之一。



手语动画合成流程

为实现合成的更高可控性，我们利用Unity3D搭建了一个手语动画合成平台。该平台输入为一段文本，再将文本按文本切分算法，结合预定义手语词汇库切分为若干词汇；根据切分结果调用手语片段数据库；最后根据手语各手语动画片段的起始段和结束段长度计算融合参数，结合融合参数建立动画状态机，输出连续手语动画合成结果依次融合各片段过渡部分得到完整的手语语句输出。



图 手语动画合成平台界面

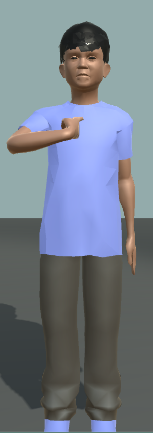


图 手语语句动画合成结果采样示例