English Research Express@NCKU Volume 3 Issue 2 - January 18, 2008
[http://research.ncku.edu.tw/re/articles/e/20080118/1.html]

Movement Epenthesis Generation for Video Synthesis of Sign Language

Ze-Jing Chuang, Chung-Hsien Wu*, Wei-Sheng Chen

Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, R.O.C.
Email: chwu@csie.ncku.edu.tw

IEEE Trans. Circuit and Systems for Video Technology, Vol. 16, No. 11, November 2006, pp1313~1323.

Individuals with hearing/speech impairment generally have problems in communication skill learning. Since deaf people have difficulty in speaking, many communication-aided approaches, such as sign language, finger spelling, lip-reading, and total communication, have been proposed to enhance their language and communication skills. Because of the impairment of vocal communication, sign-language is still the main communication method between hearing-impaired people. Although the hearing-impaired people can learn sign language with several kinds of assistance, such as books, photographs, and videotapes, none of them can provide a flexible and realistic access of sign language. Accordingly, computerized assistive learning system is proposed for sign language learning. This paper proposes a method that is able to generate a video segment of movement epenthesis using the original video clips of a real signer. Figure 1 shows the diagram of the proposed approach. The basic idea of this approach is to find the best concatenation point between two consecutive sign videos. Based on these two points, this approach can generate the smoothest epenthesis video frames to provide the sign video output.

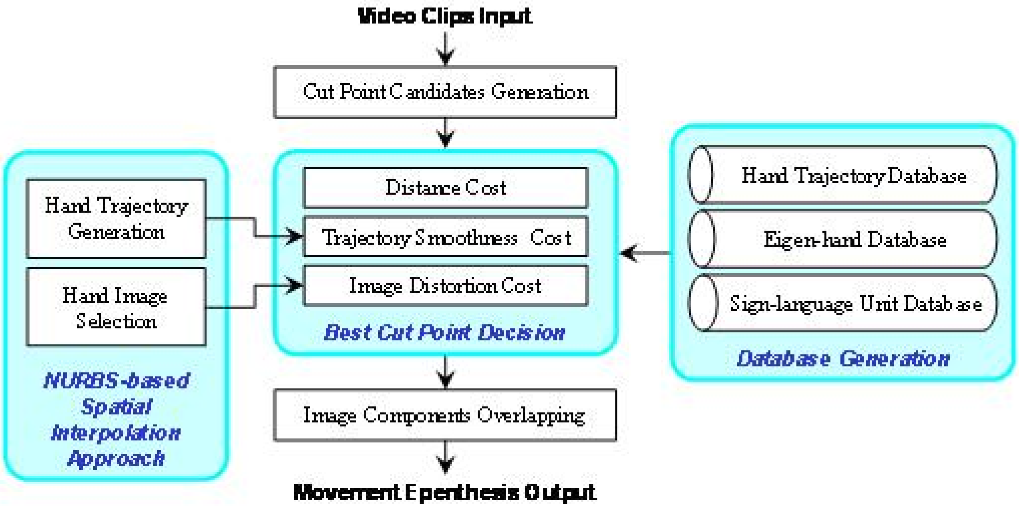
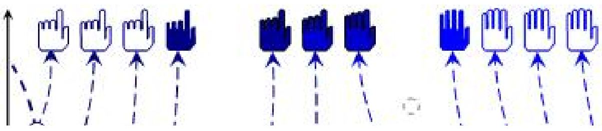

Video Clips Input
Cut Point Candidates Generation
Hand Trajectory Generation
Hand Image Selection
NURBS-based Spatial Interpolation Approach
Distance Cost
Trajectory Smoothness Cost
Image Distortion Cost
Best Cut Point Decision
Hand Trajectory Database
Eigen-hand Database
Sign-language Unit Database
Database Generation
Image Components Overlapping
Movement Epenthesis Output

Figure 1. Diagram of video-based sign-language synthesis system.

The first step for generating the movement epenthesis decides where the previous video should stop and where the next video should start, which are called “cut points.”

The best cut point pair is selected from the candidates according to the concatenation cost,



which is a linear combination of distance cost, smoothness cost, and image distortion cost. The distance cost is a normalized Euclidian distance between the hand locations at two cut points. The Non-Uniform Rational B-Spline (NURBS) function is used to generate the hand movement trajectory of movement epenthesis. In image distortion cost calculation, the NURBS function is used to generate a hand image change path for the purpose of hand image selection. From the annotation in sign-language database, all hand images in pre-captured video clips are already extracted. Therefore the hand images selection process is done with the NURBS curve. An illustration of hand image selection is shown in Figure 2.

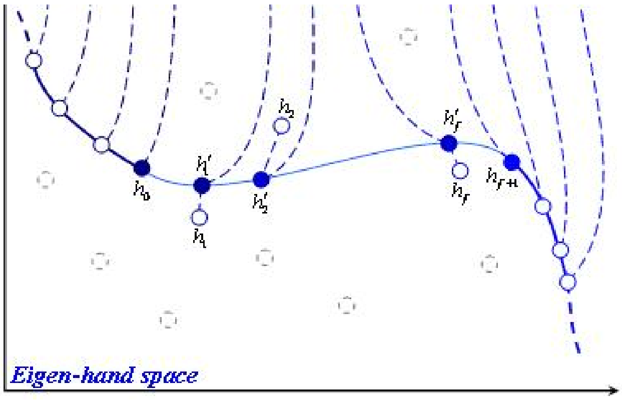
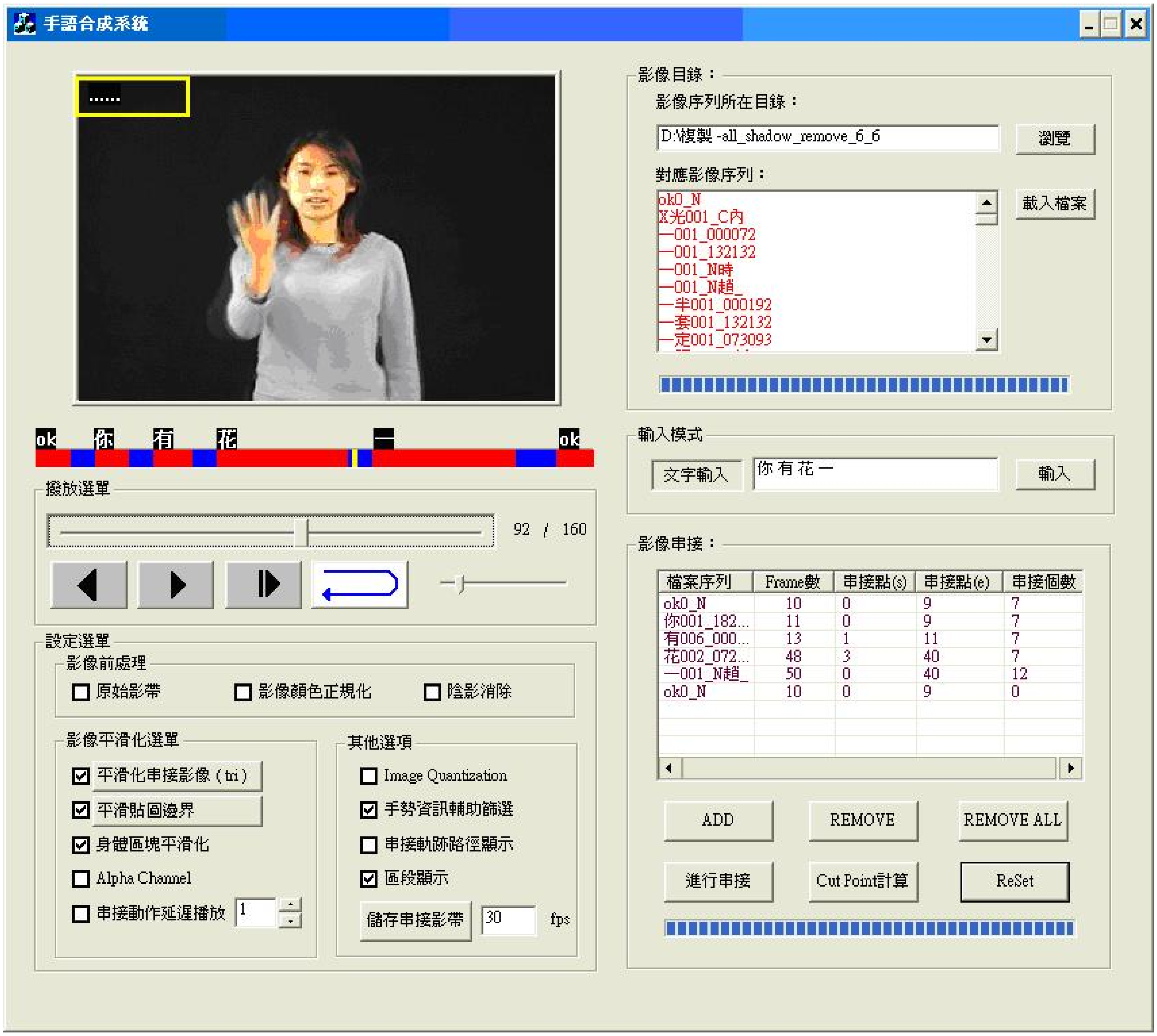


Figure 2. An illustration of the nearest hand image selection.

To support the approach, a set of sign-language databases are collected and preprocessed with image calibration, content annotation, and principle component analysis. Finally, to construct a sign-language synthesis system for demonstration and evaluation, an image component overlapping procedure is also applied as the post-process for generating an output continuous sign-language sentence video.

Using the proposed approach, a video-based Taiwanese sigh-language synthesis system was developed for evaluation on a personal computer. The final integrated system interface is shown in Figure 3. With this interface, user is allowed to input a natural language sentence. All the test video clips were evaluated by 10 normal people and 5 hearing-impaired people using the scores of fluency, understanding, and similarity. The evaluation result demonstrated that the generated movement epenthesis is not only smoothly and naturally, but also similar to the real sign-language sentence video. The result for using the combined concatenation cost also outperforms that using only one of distance cost, smoothness cost, or image distortion cost in both objective and subjective evaluations.



影像目錄：

影像序列所在目錄：

D:\複製 -all_shadow_remove_6_6

對應影像序列：

ok0_N
X光001_C內
一001_000072
一001_132132
一001_N時
一001_N趙_
一半001_000192
一套001_132132
一定001_073093

輸入模式

文字輸入: 你 有 花 一

影像串接：

檔案序列	Frame數	串接點(s)	串接點(e)	串接個數
ok0_N	10	0	9	7
你001_182...	11	0	9	7
有006_000...	13	1	11	7
花002_072...	48	3	40	7
一001_N趙_	50	0	40	12
ok0_N	10	0	9	0

設定選單

影像前處理

□ 原始影帶　□ 影像顏色正規化　□ 陰影消除

影像平滑化選單

☑ 平滑化串接影像 (tri)

☑ 平滑貼圖邊界

☑ 身體區塊平滑化

□ Alpha Channel

□ 串接動作延遲播放 1

其他選項

□ Image Quantization

☑ 手勢資訊輔助篩選

□ 串接軌跡路徑顯示

☑ 區段顯示

儲存串接影帶 30 fps

Figure 3. Interface of the constructed sign-language synthesis system.