The NIST Data Flow System II: A Standardized Interface for Distributed Multimedia Applications

A. Fillinger, L. Diduch, I. Hamchi, M. Hoarau, S. Degré, V. Stanford
National Institute of Standards and Technology
{antoine, ldiduch, ihamchi, mhoarau, sdegre, vstanford}@nist.gov

Abstract

Modern multimedia applications use an increasing number of sensors including cameras, microphones and microphone arrays. These applications must acquire and process data from sensors in real-time, which is usually beyond the capabilities of single machines. We present our distributed sensor data transport middleware, the NIST Data Flow System II, which offers network transparent services for data acquisition and processing across a local network of computers. An application is thus represented as a data flow graph, with streaming media flowing between the different computational components. This is highlighted by presenting a multimedia application tracking persons using a sensor fusion of audio and video streams.

1. The NIST Data Flow System II

The NIST Data Flow System II (NDFS II) is a middleware for sensor data acquisition, transport and distributed processing. It was originally developed to support research on pervasive environments such as smart spaces. It also facilitates the development of complex multimedia applications with very high bandwidth requirements.

The data flow system is dynamic, so computing nodes can join or leave the network at any time. Its decentralized architecture has proven to be robust under load, and is more fault-tolerant than the version one. Its cross-platform capabilities allow it to run on Linux, Microsoft Windows XP and Mac OS X, and data can be transported between these operating systems.

The version II system was developed in C++ and therefore offers an object oriented API. Its binding language capabilities also allow development of computational or data acquisition nodes in Java.

Data transport is optimized for high performance within a host using shared memory and between multiple hosts via TCP/IP wrapped in the Adaptive Communication Environment (ACE) interface[1].

Utilities, such as the Control Center, are provided to ease the use of the system, manage the complex multi node interactions, and operational status of the multimodal environments. The Control Center presents a GUI that allows creation and management of NDFS II applications. It gives the user the ability to build an application graph, i.e. to connect client nodes together and to associate each of them to a particular host. Once the distributed application is created, it can be controlled from a single computer using the Control Center from either a web browser, or a desktop of Windows, OS X, or Linux. A whole application graph can be started, stopped, and monitored, and individual node/client faults displayed, processed, and remedied at the Control Center in real-time.

2. 2. The Multimedia Demonstration

We demonstrate the capabilities of the data flow by presenting a multimedia person tracking application. This demonstration runs on several laptop computers and acquires and processes the data from two sensors: a video camera and a 64 channel microphone array[2]. The application is represented by a distributed graph containing video and audio pipelines merging in a final fusion step as shown on Figure 1. The pipelines are composed of several computing nodes, each implementing a specific functionality, and the flows to connect them.

Tracking is done in two ways: on one hand, the application acquires video data from the camera to perform face tracking of persons in the field of view, and on the other hand multichannel audio data is used to track the dominant sound source paths; which are assumed to be the speaker paths.

In order to track sound source paths, we first capture multichannel audio data from the microphone array (*Capture_Audio_Array* client node), then we

This paper is authored by an employee(s) of the United States Government and is in the public domain.

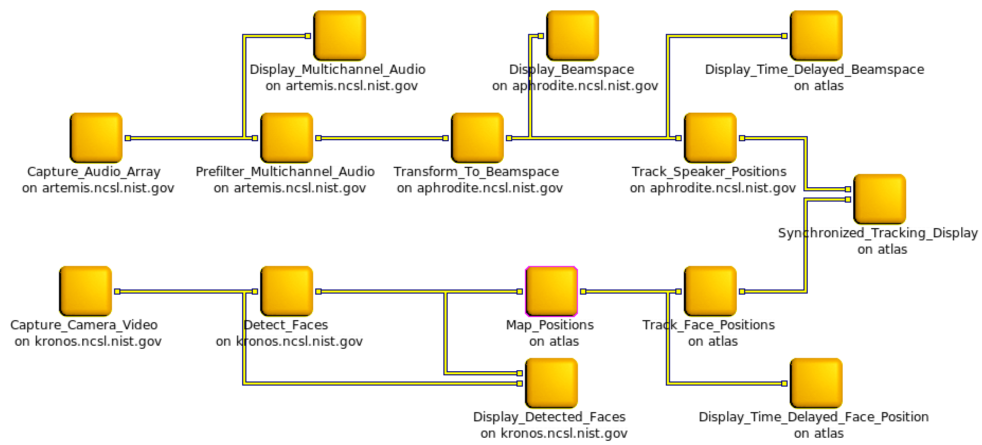

Figure 1. The multimedia application graph. This graph uses distributed systems to acquire multichannel audio, and video, data to localize and track faces, and speakers as they move.

process this data with a preemphasis and FIR bandpass filters (*Prefilter_Multichannel_Audio*).

Next, a delay-and-sum beamforming algorithm, *Transform_To_Beamspace,* computes energy as a function of angle of incidence to provide directions of the sound sources. Using time averaged direction histories we track the dominant sound source path by applying a multi stage decision line tracking algorithm[3], *Track_Speaker_Positions*.

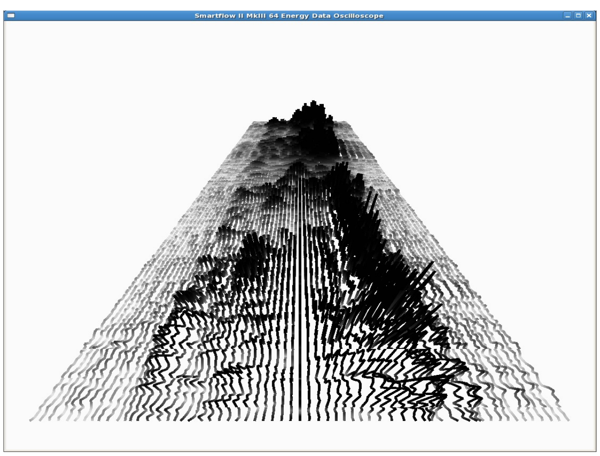

Figure 2. View of sound direction history in beamspace from the multichannel audio.

In the video pipeline, the captured video stream (*Capture_Camera_Video*) is used to detect faces applying the well known Viola-Jones algorithm (*Detect_Faces*). In order to track the face positions, the history of the face positions over time are mapped to a direction (*Map_Positions*) and processed using the same multi stage decision line tracking algorithm (*Track_Face_Position*) used in the audio pipeline.

In a last step, the tracking results from the audio and video pipelines are fused and displayed *(Synchronized_Tracking_Display*) as seen at right in Figure 1. By correlating the results of the two pipelines it is then possible to identify active speakers.

The two pipelines, which have very different sampling and data-rates, can be fused either synchronously or asynchronously. In this demonstration we use the NDFS II *ad hoc* stream synchronization capabilities to accomplish the necessary time coherent fusion.

To monitor the internal states of the processing pipelines, several displays are connected at different points of the data flow. We represent raw data in *Display-Multichannel-Audio* (Figure 5); beaformed data in *Display-Beamspace* (Figure 3); beamformed delayed data in *Display-Time-Delayed-Beamspace* (Figure 2); face position and video data *Display-Detected-Faces* (Figure 4 at right) and face position history *Displayed-Time-Delayed-Face-Position* at left.

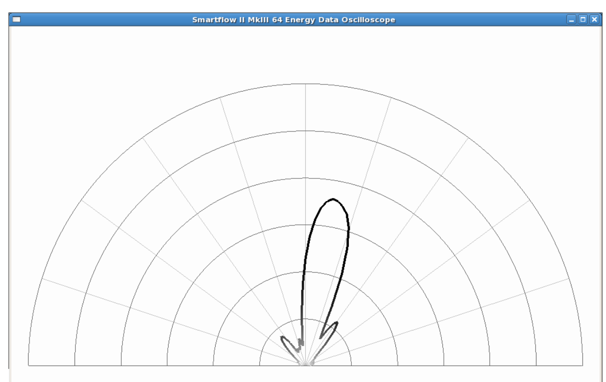

Figure 3. Polar coordinate view of current sound direction from the NIST Mk-III Array.

This paper is authored by an employee(s) of the United States Government and is in the public domain.

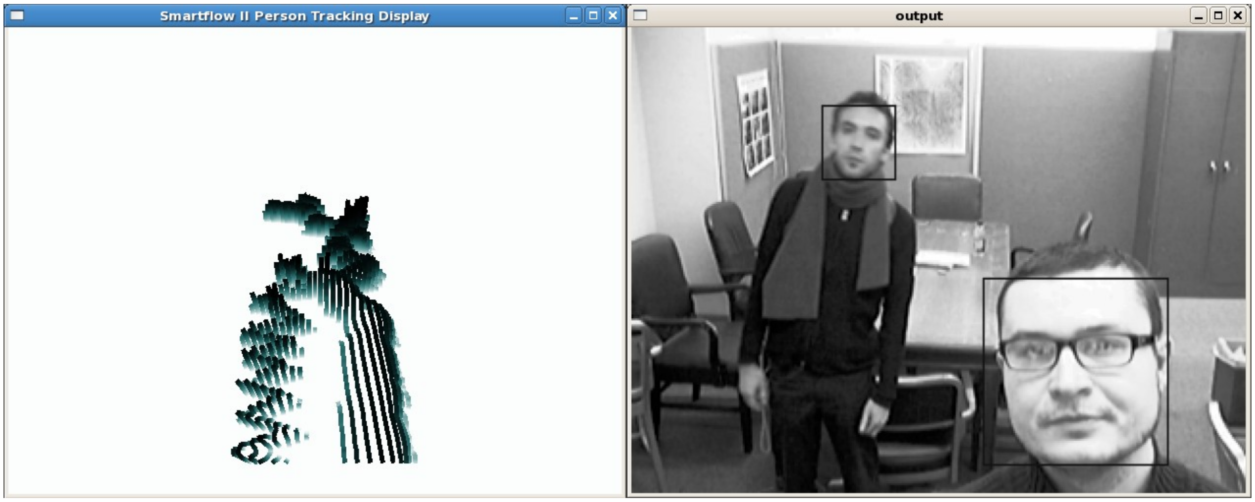

Figure 4. The face position track history as a function of time(left), and an example video frame with the current regions of interest marked bounding boxes (right).

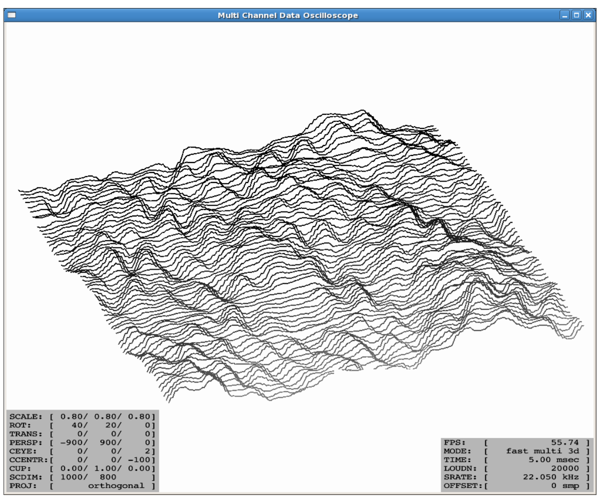

Figure 5. View of the 64 audio channels from the NIST Mk-III array used to compute the beamspace representation in Figure 2.

3. Disclaimer and Statements

The NIST Data Flow System II software (NDFS II) was developed at the National Institute of Standards and Technology (NIST) by employees of the Federal Government in the course of their official duties. Pursuant to title 17 Section 105 of the United States Code this software is not subject to copyright protection and is in the public domain.

Certain commercial products may be identified in order to adequately specify or describe the subject matter of this work. This implies no recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that such products are necessarily the best available for the purpose.

The Data Flow is an experimental system. NIST assumes no responsibility whatsoever for its use by other parties, and makes no guarantees, expressed or implied, about its fitness for any particular purpose.

The National Institute of Standards and Technology and the Smart Spaces Project would appreciate acknowledgments if the tools are used.

4. References

[1] D. Schmidt, *The ADAPTIVE Communication Environment - An Object-Oriented Network Programming Toolkit for Developing Communication Software,* Sun Users Group, and http://www.cs.wustl.edu/~schmidt/PDF/SUG-94.pdf, 1994**.**

[2] Vincent M. Stanford, John S. Garofolo, Olivier Galibert, Martial Michel, Christopher Laprun, *The NIST smart space and meeting room projects: signals, acquisition, annotation, and metrics*, Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, April 2003, Vol. **4**, pages 6-10.

[3] J-C. Di Martino, J.P. Haton, and A. Laporte, "Lofargram Line Tracking by Multistage Decision Process", *Acoustics, Speech, and Signal Processing,* ICASSP-93, Minneapolis, MN, USA, 27-30 Apr 1993, pp. 317-320.

This paper is authored by an employee(s) of the United States Government and is in the public domain.