Improving Registration Using Active Shape Models and Depth

Colin P. Bellmore, Roxanne L. Canosa*
Rochester Institute of Technology, Rochester New York

Figure 1: *Far left: candidate registration points – RGB image points shown in green and depth image points shown in red. Left middle: transparent overlay of RGB and depth images with depth window border. Center: active shape model fitted to RGB image. Right middle: active shape model fitted to depth image. Far right: registration results after deformation – RGB points in green and depth points in red.*

1 Introduction

The problem of registering multiple images taken of the same scene is particularly difficult in domains where different sensors are used to collect the information. Registration techniques are used to combine the independent information into a single image for further processing or for additional insight into the data. In video, sequential images are usually highly correlated, making registration simpler; the exception is when a new scene enters the viewfinder or when extreme camera movement results in large differences between sequential images. In these cases, far less information is available for registration. Using multiple sensors trained on the same scene is helpful; however, the precise relationship between the sensors must be obtained to compare the data spatially as well as temporally. In these cases, landmarks are often useful.

Registering RGB webcam images with depth images is difficult due to the absence of detectable landmarks in the depth channel. Usually, an additional motion or shading model is imposed to correlate "depth" pixels to color pixels, taking advantage of the fact that an object's depth and its color are spatially well-registered in the real world. In addition, registering a depth sensor with an RGB webcam requires finding and modeling both cameras' parameters as well as characterizing the noise from each sensor. We propose using a non-rigid deformation technique based on active shape models (ASMs) [Cootes and Taylor 1992] to circumvent the need for explicitly modeling camera parameters. An ASM is a statistical shape model that can provide information for overcoming the difficulties inherent in registering image patches obtained from color and depth sensors. The ASM is used in the absence of real landmarks and works by assesing curvature and learning discrimintating points on a per-channel basis. This technique was applied to real-time video and tested on ASMs trained from low-resolution facial images generated from an inexpensive commercially available RGB+D sensor. The results show that the deformation technique improves registration under specific conditions.

2 Our Approach

Our system used an active shape model for registering images originating from separate depth and color channels. This required that object points located in the world map to the corresponding pixels in each image. To achieve this, we first geometrically aligned the depth and color sensors and then spatially rectified the two image streams. Additionally, the depth data was filtered to reduce noise and then the face was fitted to both channels for the purpose of locating control points for the ASM. The rectified and non-rigidly deformed depth map was then overlaid onto the RGB image, producing a registered 4-channel image.

The overall approach consisted of five processing stages: (1) localization of the face using a motion-based depth silhoutte, (2) projective viewpoint transformation to model the relationship between the RGB and depth images, (3) motion-restricted median filtering to improve the accuracy of the depth measurements, (4) active shape model training and fitting, and (5) deformation of the ASM control points. For the final deformation step, local image registration was required because certain areas of the image needed different amounts of deformation. Since the non-linear nature of the deformation can cause computational efficiency problems, a locally weighted mean field smoothed by a Gaussian allowed the model to spend processing resources on high areas of deformation while not affecting previously well-registered areas.

The results after deformation were analyzed for validity by measuring the distance between the transformed points and the control points of the ASM. The results were evaluated by comparing our 14-point ASM model to previously published 58-point and 68-point models under conditions of varying lighting and facial structures. Evaluation metrics included mean fitting time, mean errors, and standard deviation of errors. Results showed that our model has lower mean fitting time as compared to state-of-the-art 58-point and 68-point models (41 msec versus 428 msec and 544 msec respectively) and lower mean error (2 pixels versus 7 pixels and 8 pixels respectively), although the lower resolution of our test images compared to the other models' test images contributed to our smaller average error. Overall, this study showed that ASMs can be successfully used for registration of color and depth images.

References

Cootes, T., and Taylor, C. 1992. Active shape model – smart snakes. In *Proceedings of the British Machine Vision Conference*, 266275, 1829–1841.

*e-mail:cpb8010@rit.edu

SIGGRAPH 2012, Los Angeles, California, August 5 – 9, 2012.
ISBN 978-1-4503-1435-0/12/0008