# Non-planar Infrared-Visible Registration for Uncalibrated Stereo Pairs Supplementary Material

Dinh-Luan Nguyen
Faculty of Information Technology
University of Science, VNU-HCMC
Ho Chi Minh city, Vietnam
1212223@student.hcmus.edu.vn

Pierre-Luc St-Charles, Guillaume-Alexandre Bilodeau
LITIV lab., Dept. of Computer & Software Eng.
Ecole Polytechnique de Montreal
Montreal, QC, Canada
{pierre-luc.st-charles,gabilodeau}@polymtl.ca

## 1. Visual results for planar scenes

In this section, we visualize the registration results for planar scenes. For better comparisons between our proposed framework and those of St-Charles *et al*. [2], Sonn *et al*. [1], we show three video scenes at the beginning, middle, and end of each LITIV video [3] to demonstrate the temporal registration process accuracy. The first column in each figure shows the thermal infrared image of the scene, and the other columns show pairs of image and polygon registration results for our approach, St-Charles *et al*.'s and Sonn *et al*.'s respectively.

As it can be seen from Figure 1 (LITIV 1, 2, 3), Figure 2 (LITIV 4, 5, 6), and Figure 3 (LITIV 7, 8, 9), our proposed framework achieves good registration in the early frames of videos. More specifically, we obtain low polygon registration error (gray area) in the beginning frames at all LITIV videos except LITIV 3 and LITIV 9. However, the proposed framework quickly obtains the right registration by applying our video rectification strategy. In contrast, St-Charles *et al*. [2] and Sonn *et al*. [1] still get high registration error in the early stages. However, although their works achieve low registration error later in most sequence, these errors are still higher than ours. Furthermore, our robustness is also revealed by doing experiments on LITIV 6. Because people in this scene are very small, it is difficult to get features. While the other methods get acceptable registration errors in the beginning, they fluctuate their final registration matrices so that noise dramatically affects their results.

## 2. Visual results for non-planar scenes

In Figure 4, we visualize the proposed framework's registration in our new dataset. The column shows the final registration (derectified images) of Video 1, 2, 3, and 4. Since each video is captured from different camera setups, the rectification result for each video is unique. As it can be observed from Figure 4, the proposed framework can handle nearly all situations. However, there are still some scenes, e.g. Video 1, which our framework does not register successfully since there are so many people in one frame. Besides, the reason for why blob final registration is not smooth is that we use the invert matrix $H_2$ to derectify scenes. This matrix is created from fundamental matrix so that multiply with ${H_2}^{-1}$ can only approximate the result, not the exact stage as raw input video. Hence, final registration blobs seem not smooth.

In general, our proposed framework successfully register non-planar TIR-Vis video.

## References

[1] S. Sonn, G.-A. Bilodeau, and P. Galinier. Fast and accurate registration of visible and infrared videos. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 308–313. IEEE, 2013. 1, 2, 3, 4

[2] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. Online multimodal video registration based on shape matching. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on*. IEEE, 2015. 1, 2, 3, 4

[3] A. Torabi, G. Massé, and G.-A. Bilodeau. An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications. *Computer Vision and Image Understanding*, 116(2):210–221, 2012. 1

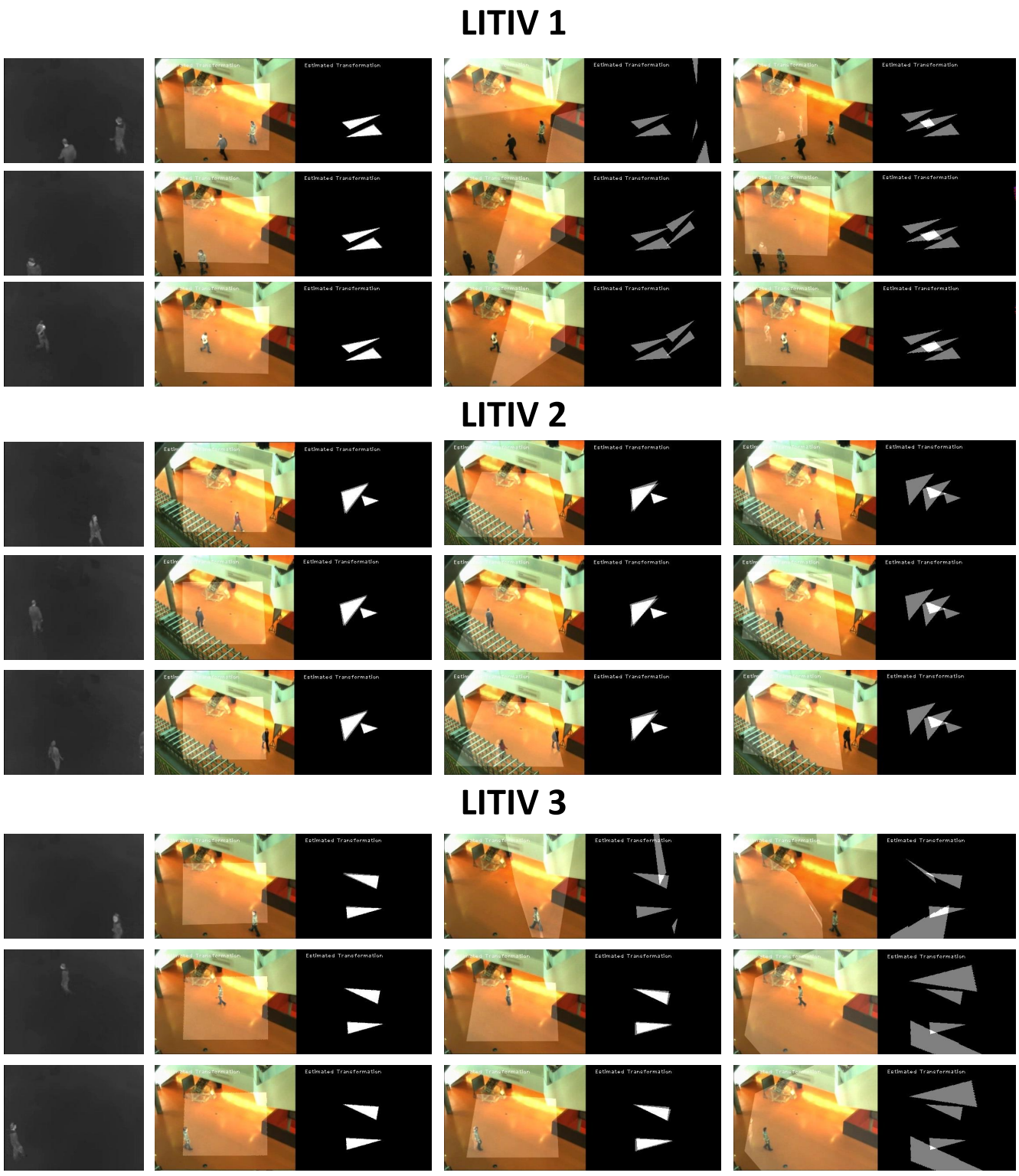


Figure 1. Visual registration comparison between the proposed framework ($2^{nd}$ column), St-Charles *et al*. [2] ($3^{rd}$ column), and Sonn *et al*. [1] ($4^{th}$ column) for dataset LITIV 1, 2, and 3.

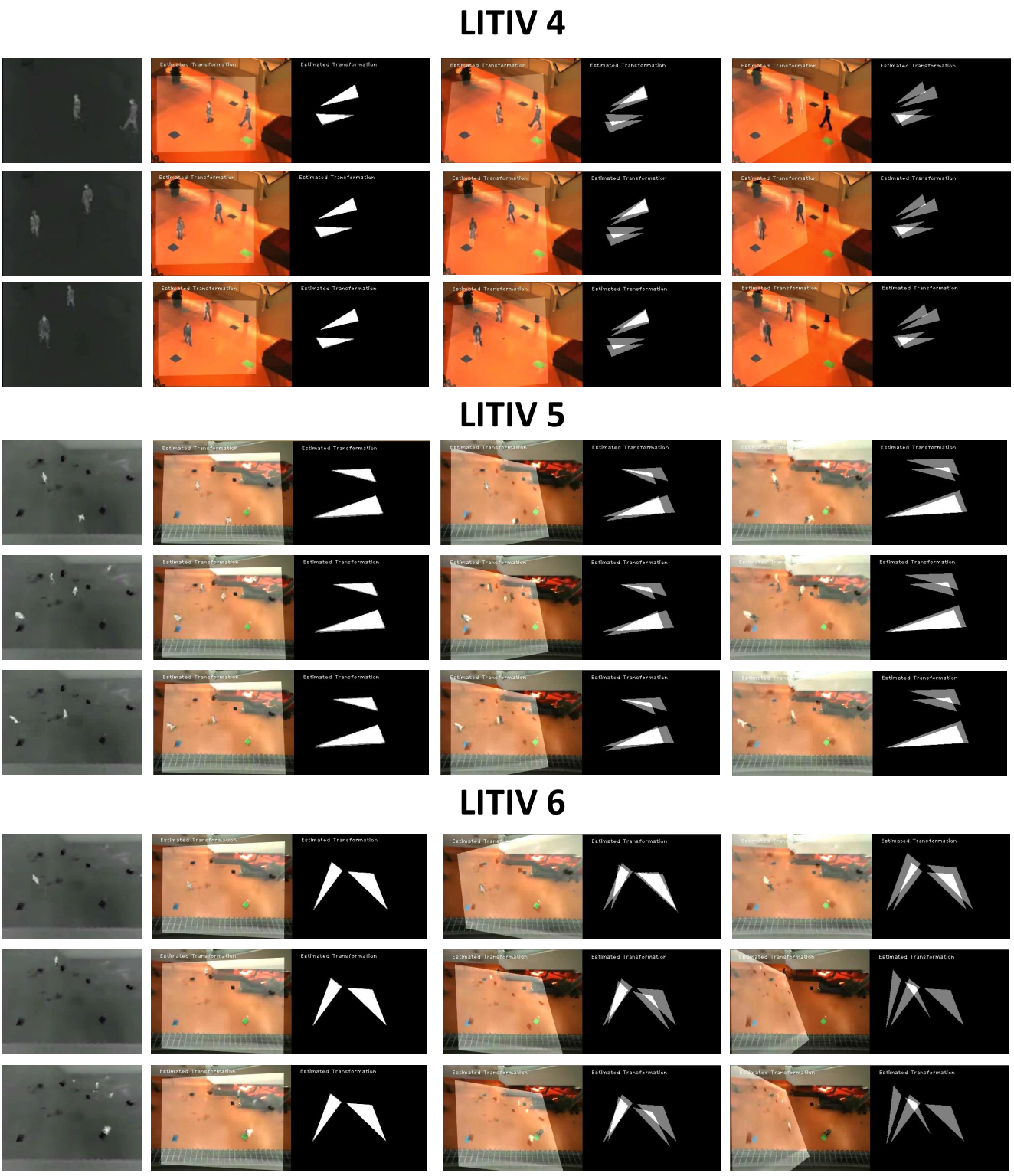


Figure 2. Visual registration comparison between the proposed framework ($2^{nd}$ column), St-Charles *et al*. [2] ($3^{rd}$ column), and Sonn *et al*. [1] ($4^{th}$ column) for dataset LITIV 4, 5, and 6.

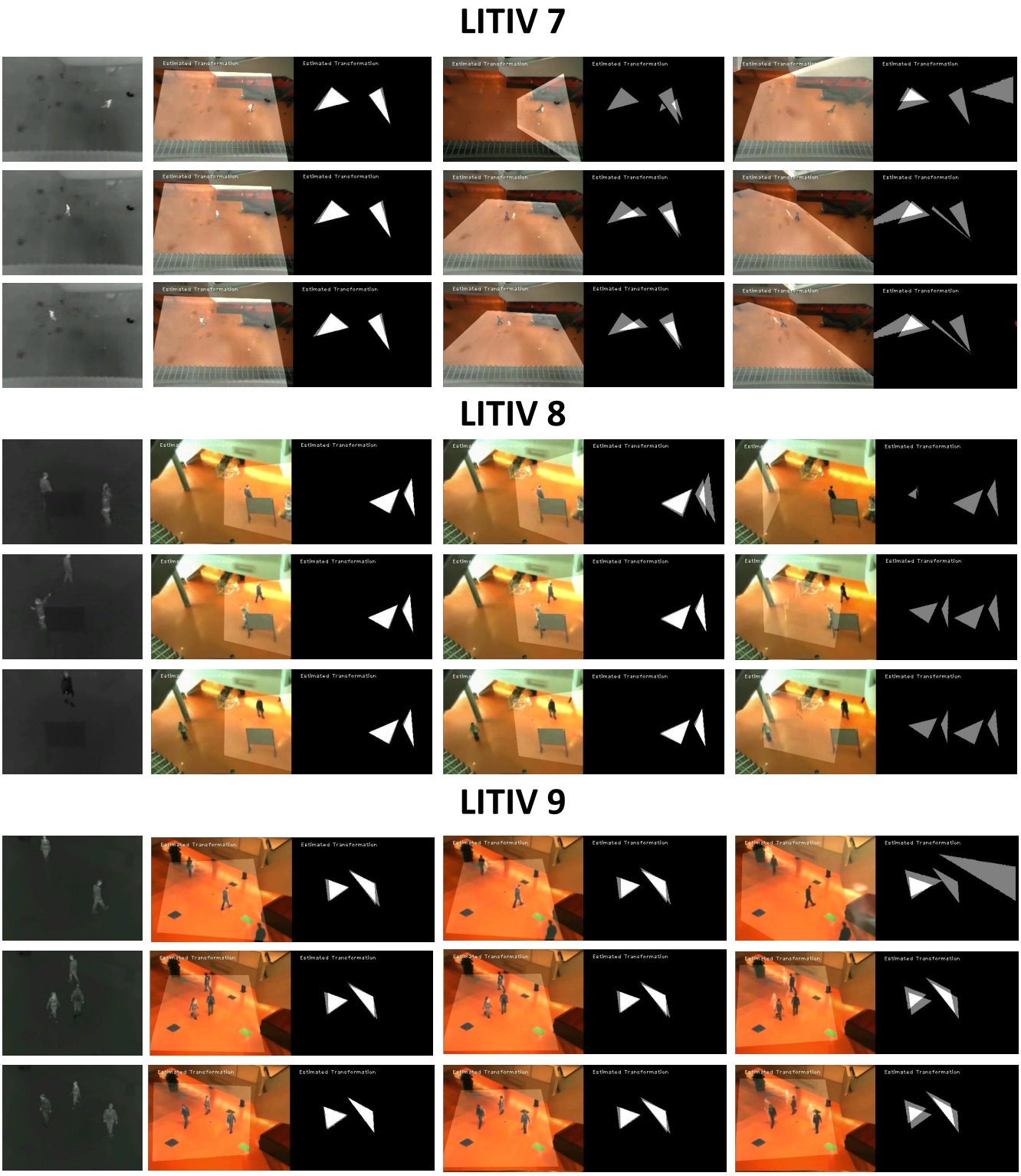


Figure 3. Visual registration comparison between the proposed framework ($2^{nd}$ column), St-Charles *et al*. [2] ($3^{rd}$ column), and Sonn *et al*. [1] ($4^{th}$ column) for dataset LITIV 7, 8, and 9.

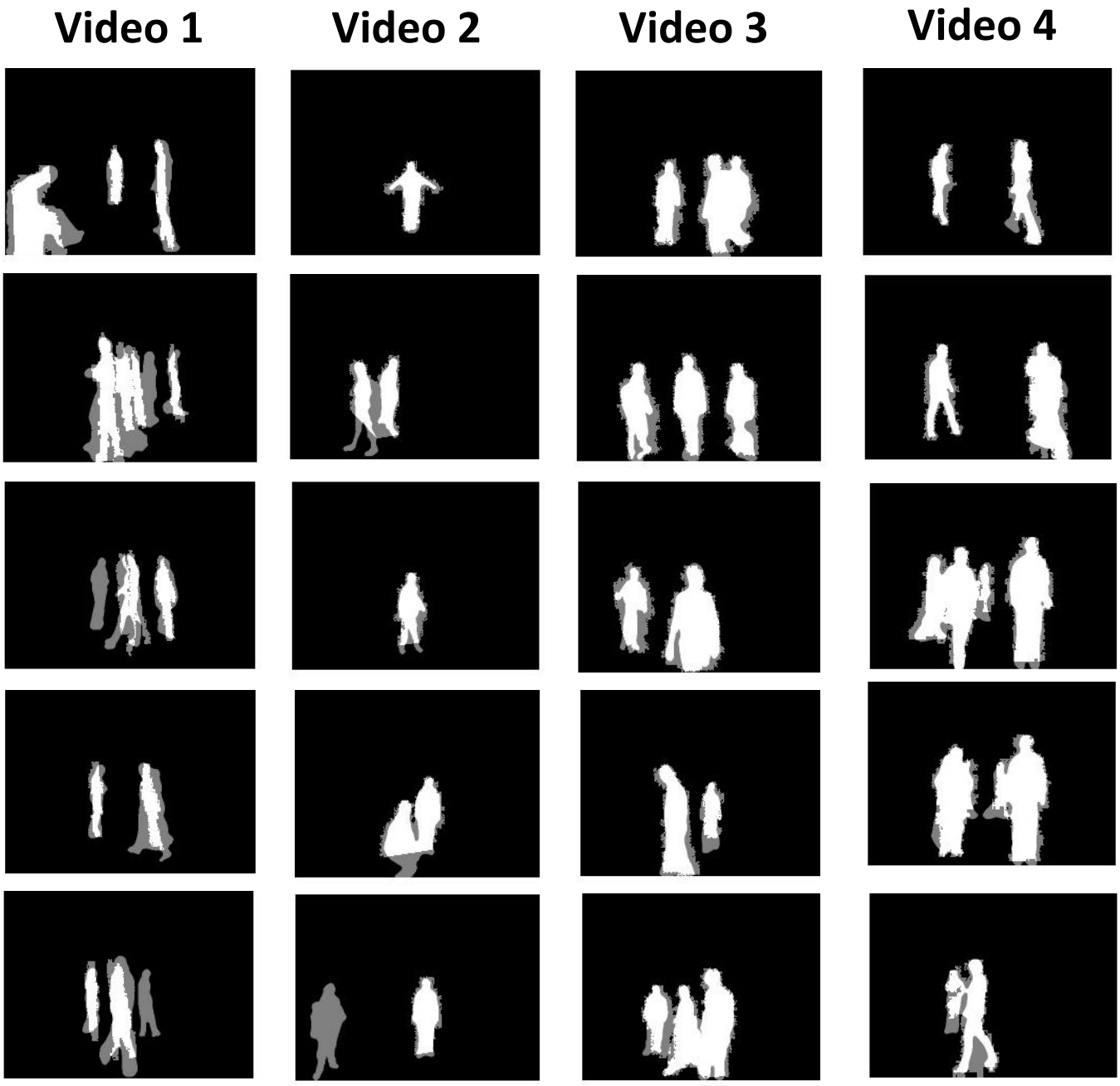


Figure 4. Visual registration comparison in non-planar scenes. In each column, we show the final registration for raw video scenes.