False Positive Outliers Rejection for Improving Image Registration Accuracy
Application to Road Traffic Aerial Sequences

Ines Hadj Mtir[1], Khaled Kaâniche[1], Pascal Vasseur[2] and Mohamed Chtourou[1]
[1]*Intelligent Control, Design and Optimization of Complex Systems, ENIS, University of Sfax, Sfax, Tunisa*
[2]*Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes, LITIS, University of Rouen, Rouen, France*

Keywords: Outliers Rejection, Features Matching, Image Registration, Motion Compensation, Vehicles Detection, Aerial Sequences.

Abstract: The objective of our system is to detect vehicles from aerial sequences. Theses sequences are taken from a camera mounted on UAV which flies over roads and highways. Our approach is to firstly compensate the motion introduced by the dynamic behaviour of the camera. This leads us to a problem of image registration. The moving regions (vehicles) are after that extracted using residual motion. The aim of this paper is to present a combined method for features matching and outliers rejection to increase the accuracy of the registration phase. We use first, the SIFT descriptors and then outliers are rejected using geometric constraints. This leads to a better registration and a minimum of false alarms in the detection phase.

1 INTRODUCTION

Image registration is widely used in remote sensing, cartography, medical image registration, image mosaicing computer vision application and pattern recognition (Zitova and Flusser, 2003). Multi-modal image registration (brain CT/MRI images or whole body PET/CT images) is mostly used for medical application to obtain a more complex and detailed scene or to follow the evolution of a tumor. Viola and Wells 1997 uses mutual information as a criterion to register medical images using gradient descent optimization method. An overview of medical image registration techniques can be found in (Whawahre et al., 2009).

Template registration is used to localize a template in the scene or to register aerial or satellite images to GIS map (Nakagawasai and Saji, 2011). Another application of image alignment is multi-view point registration which aims to panorama and mosaic construction (Kang and Ma, 2011). In their paper (Vivet et al., 2011) proposed a new mosaic creation method named direct local indirect global registration (DLIG), the registration is iteratively computed by sequentially imposing a good local match and global spatial coherence. They compared their DLIG method to frame to frame and frame to mosaic registration method and proved better performance in reducing the accumulation error problem. A tutorial on image alignment and stitching has been proposed in (Szeliski, 2006). Azzari, 2007, proposed a real time image mosaicing method which is divided on to step: Frame to Frame registration, using SIFT features and RANSAC to eliminate outliers, and a Frame to Mosaic registration to refine the result and eliminate photometric misalignment, using a histogram specification approach.

To detect changes or moving object in the scene, multi-temporal registration is needed, it uses different images taken from the same scene but taken at different time. In our application we have at the same time a change in the point of view, as the camera is mounted on an unmanned aerial vehicle (UAV), and also a multi-temporal image capture. So before the detection of the moving objects (vehicles), a phase of dominant motion compensation is needed. A geometric transformation has to be determined to estimate the transformation between a reference frame I0 at time t and the target one I1 at time t+1. Once consecutives images are registered, the detection of moving objects is intuitively obtained by residual motion estimated from the optical flow field deduced from the image Brightness Constancy Equation (IBCE) (Medioni et al., 2001).

Hadj Mtir I., Kaâniche K., Vasseur P. and Chtourou M..
False Positive Outliers Rejection for Improving Image Registration Accuracy - Application to Road Traffic Aerial Sequences.
DOI: 10.5220/0004038102740279
In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2012), pages 274-279
ISBN: 978-989-8565-22-8

The purpose of our algorithm is to reduce the incorrect matches rate and improve the accuracy of the registration phase. Our scenes contain some moving objects but the transformation has to estimate the motion of the image background. The existing methods of feature matching try to reject outliers. But, in our case, we need to reject false positives matches which are detected on the moving objects. This can biases the estimation of the background motion. So we introduce a geometric criterion to eliminate false positive and false negative matches which are respectively: incorrect matched points and correct matched points attached to moving objects.

The paper is divided as follows: in section 2 we have an overview of global and local image registration techniques. In section 3 we present a formulation of the registration problem. A description of the traditional image registration algorithm and our proposed amelioration which adds geometric criterion are explained in section 4. Section 5 presents some experimental results.

2 PREVIOUS WORKS

Image registration can be approached with global or local method. The global one consists in optimizing a certain criterion until obtaining a geometric transformation which fits correctly the two registered frames. Used criterion is usually the sum of squared difference (SSD) of the whole image luminance, the correlation or the mutual information…etc. (Whawahre et al., 2009). These methods need textured surface and are very time consuming since they work on the total number of pixels of the image. They can also be sensitive to the luminosity change of the image (SSD), can not handle a very large rotation, translation or scale changes and can easily fall into a local minimum.

Local methods are usually divided into four steps: feature detection, feature matching, transformation function estimation and image re-sampling. A review of image registration approaches can be find in (Zitova and Flusser, 2003; Xiong and Zhang, 2009a; Xiong and Zhang, 2010). Features can be edges, corners, lines, regions or a combination between them. These methods are less consuming time as they work with some relevant and reliable part of the image. Many features points have been proposed and improved all over the time: Moravec, Harris and Stephens, Trajkovic, SUSAN detector.

Every time the detectors try to be less sensitive to noise, invariant to affine transformation and rotation or scale changes. Laplacian of Gaussian and Difference of Gaussian are invariant scale blobs detector, on which, is based the most known and robust features detector: SIFT (Scale invariant feature transform) (Lowe, 2004). Govender, 2009 showed that SIFT is one of the best distinctive detector. The requirements of a feature detector are: Every “true point” must be detected; No false alarm must be detected; Points must be well localized, Detector must have a high repeatability rate (stable between different images); Detector must be insensitive to noise, Invariant to rotation and scale changes and finally; Detector must have a reduced algorithmic complexity for the real time application.

Once the features are detected, the next step is feature matching. Matching can be done using a similarity criterion between two window centred on the feature point like SSD, NCC (normalized correlation), Mutual information…etc. But this window is only adapted for distortion caused by a translation. Those similarity criterion can also be sensitive to noise and illumination change, as well that they need textured regions.

Invariant descriptors are well adapted to describe a feature point. Schmid, 1992 proposed an eight components descriptor based on different derivatives order of the luminance function. Lowe proposed, in addition to the features detector SIFT, a 128 elements descriptor estimated from gradient oriented histograms. Many other variant of the SIFT descriptors has been also proposed: RIFT (Lazebnik et al., 2004), PCA-SIFT (Ke and Sukthankar, 2004), GLOH (Mikolajczyk and Schmid, 2005), GRIFT (Sungho et al., 2006) and SURF (Bay et al., 2006). These variants have increased its robustness, distinctiveness and even reduced its descriptor size (PCA-Sift and GLOH). Mikolajczyk and Schmid, 2005 compared some local descriptors and proved that SIFT and GLOH present the highest matching accuracy.

Feature descriptor must be: Invariant: the same point in different frame must have the same descriptor; Unique: two different points must have two different descriptors; Stable: the descriptor of the same primitive but with some scale or rotation change must be the same as the original one; Independent: if the descriptor is a vector, its elements have to be independent (generally not feasible).

3 PROBLEM FORMULATION

In our application (see figure 1) the UAV flies over roads and highways with sometime a very low altitude. So the scene is generally not textured and presents a very low number of potential feature points. Also as it is presented in figure 1; with a low altitude vehicles take an important part from the image information. And many features points are concentrated on these moving objects. Or we aim to estimate a geometric transformation of the background between two consecutive frames. Traditional detectors, SIFT and its variants, will correctly detect and match these points. But we need to reject these false positive outliers to only estimate the background motion.

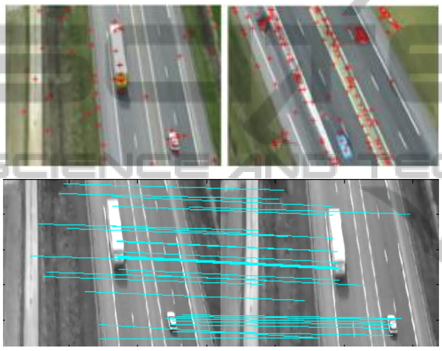


Figure 1: Top: Results of the SIFT detector. Bottom: The result of the matching process.

Many approaches have been presented in order to reject outliers. RANSAC is widely used for outliers rejection. This is an iterative method to fit a geometrical model to the dominant number of points. The accuracy of this method is inversely proportional to the percentage of the outliers. It does not work when there are more then 50% of outliers, except that, this can be usually our case: with non-textured images and with low altitude a high rate of features will be concentrated on vehicles.

Spatial relation must be added to separate the false positive and false negative outliers from the true positives ones. ICP iterative closest point (ICP) (Besl and Mckay, 1992) is another simple iterative approach to find rigid transformation but which need a good initial estimation to guarantee a convergence to the correct solution. Then probabilistic methods have been proposed to overcome this limitation like in (Luo and Hancock, 2001; Liu , 2012; Saromà et al., 2010) where they use a graph matching algorithm to integrate a spatial solution.Our application has to be a real time one: before the capture of the next frame, the system must have already detected the vehicles between the last two frames. So we need an efficient and a fast image registration step.

We propose to combine the matching phase of the SIFT algorithm with a spatial verification approach to eliminate the incorrect matched points due to the aperture problem, the non-textured environment and the repeatability of the road marks. Also to eliminate the false positive feature points above moving objects.

4 IMAGE REGISTRATION ALGORITHM

4.1 SIFT and RANSAC Algorithm

We compared the performance of our algorithm to a traditional one for image registration: keypoints are detected and matched using SIFT descriptor (figure 1) then RANSAC is applied to fit the adapted homography transform to register two successive frames (Martin et al., 1981; Brown and Lowe 2002; Azzari, 2007; Wei et al., 2008).

RANSAC is now applied to fit the best geometric transform which wrap the target image I1 to the reference one I0. From the set of keypoint matched at least 4 not collinear pairs are randomly extracted and an estimation of the homography matrix (3x3) is estimated, All other data are then tested against the estimated model and, if a point fits well to this homography, it will be considered as a hypothetical inlier.

This is repeated until a good estimated model is found and when sufficiently points have been classified as hypothetically inliers.

Homography is a 3x3 projective matrix which wrap a feature point P1 in I1 with coordinate (u_1,v_1) to its correspondent one P0 in the reference frame I0. Theoretically P0=H × P1, H is estimated with a direct linear transform DLT (1):

$$\begin{pmatrix} u_0 \\ v_0 \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & 1 \end{pmatrix} * \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} \quad (1)$$

The last step is to register the target image to the reference one, an interpolation is necessary as many pixel coordinate will not be found with the transformation $X_1 = H \times X_2$.

Figure 6 present some result of matching keypoint with this algorithm. We can often see the high rate of false positive matched point above vehicles.

This low precision of the SIFT-RANSAC

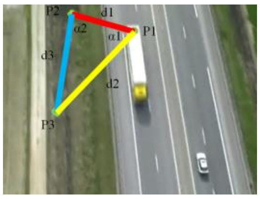


Figure 2: The triangle descriptor estimation.

algorithm will give a biased homography function and so a low accurate image registration step.

4.2 Geometric Filtering Method

As explained in the last section false negative and positive matches are not correctly rejected. We propose to add after the SIFT matching step a verification step based on spatial criterion. In fact for each matched point P1 from the reference image we take randomly 2 others not collinear points (P2, P3) from the same image (figure 2) and estimate a descriptor vector V with:

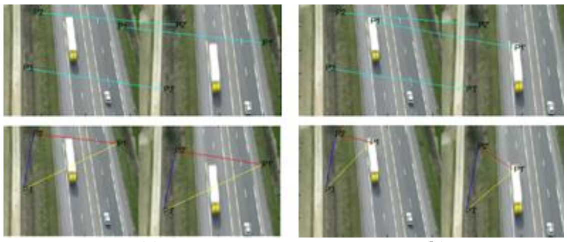
$$V = (\frac{d1}{d3}, \frac{d2}{d3}, \alpha_1, \alpha_2, C_1, C_2) \qquad (2)$$

d1, d2 and d3: are the Euclidian distance respectively between the points (P1, P2), (P1, P3) and (P2, P3). The angles are found from the cosine and sinus which are estimated from the scalar and dot product of the vectors $\overrightarrow{\mathrm{P1\,P2}}$, $\overrightarrow{\mathrm{P1\,P3}}$ and $\overrightarrow{\mathrm{P2\,P3}}$. C1 and C2 are the correlation coefficients between 2 windows (5x5) centred on the feature points.

For the correspondent points P1’, P2’ and P3’ of P1, P2 and P3 in the target image, we estimate the descriptor vector V’. The pair of points (P1, P1’) is supposed to be correctly matched if the Mahalanobis distance between the two descriptors V and V’ is under a certain threshold. For a more robust outliers rejection step we repeat this process K times (K=3), if the pair of point is at least K-1 times identified as a correct match, it will be accepted as inlier.

The advantage of this method is that if a pair of point is not correctly matched with the SIFT descriptor the triangle (P1, P2, P3) will not be equal to the correspondent one (P1’, P2’, P3’). We notice also that the triangle descriptor vector is invariant to rotation, translation and scale factor. We see in figure 3 an example of a correct matched point (figure 3.a) and a false positive matched point (figure 3.b). In our case vehicle feature points are considered as incorrectly matched, so even if P1 and P1’(features points detected on vehicles) are correctly matched with the SIFT descriptor we can reject them by taken in account spatial information.

Figure 6 shows some example of amelioration of the matching results in front of the result obtained with the SIFT+RANSAC algorithm.



(a) (b)

Figure 3: (a) A case where points are correctly matched. (b): A case where points are not correctly matched.

5 EVALUATION

To compare the performance of both algorithms a precision rate is estimate (figure 4.a):

$$1 - \text{Precision} = 1 - \frac{\text{Nb of correct matches}}{\text{Nb of total matches}} \qquad (3)$$

We took also a point P with coordinate (u, v) and estimate the coordinate of the correspondent point P’ with the homography obtained with RANSAC algorithm HR, triangle algorithm HT and a manual estimated one HM. The Euclidian distance between the point P’ found with HR and the one found with HM is estimated and compared to the result found with the HT homography (figure 4.b).

The last comparison is done with the estimation of the normalized SAD (Sum of Absolute Difference) error between the reference frame and the wrapped one obtained with the HR and then with the HT homographies. (figure 4.c). We evaluated both algorithms on 48 samples from our data base image sequences.

Figure 4 shows how our algorithm outperforms the RANSAC one. In fact it has a very high precision rate even in the case of low altitude and homogonous frames. Not only are the false matches eliminated, but also the false positive feature points.

Mosaics are also created to show the efficiency of our method. Figure 5 presents some mosaics obtained from aerial sequences. Frame to mosaic registration is used to eliminate the accumulation error. We show that mosaics present too few distortions. With an efficient image registration step, moving objects are easier to find and less false alarm are detected.

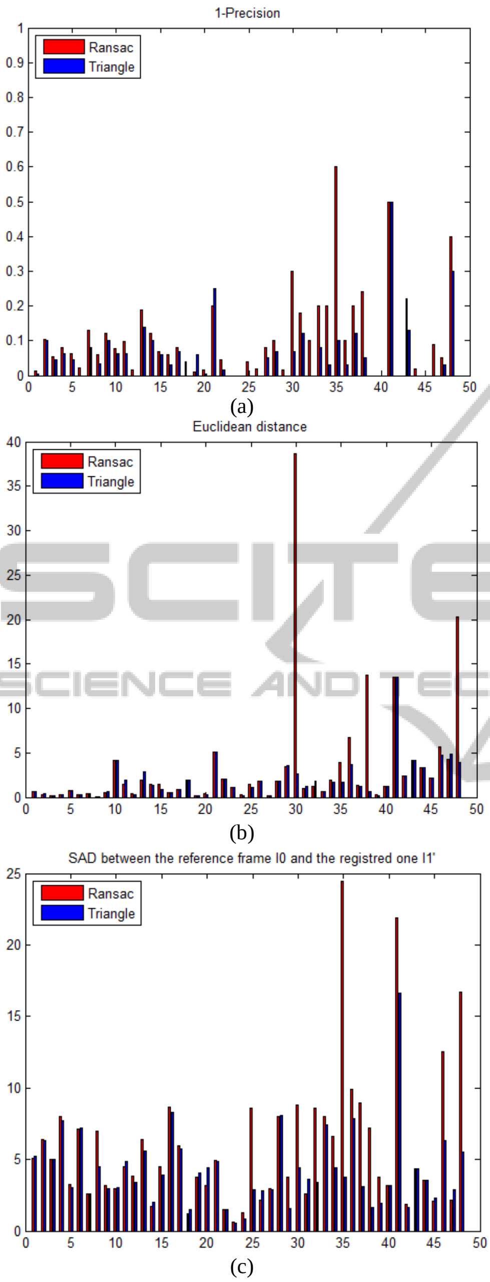


Figure 4: Evaluation. a): 1- Precision. b): Euclidian distance. c): SAD.

6 CONCLUSIONS

Image registration step is very important to assure good moving object detection. We proposed in this paper a solution to reject false negative and false positive matched point a find an optimal geometric transformation which correctly wraps a target image to a reference one. Our performance comparison showed a higher amelioration especially in the precision rate. With a low computational consuming time (in the same order than the RANSAC one, in the order of 1s) we proposed a simple and efficient solution to reject outliers.

REFERENCES

Azzari P., 2007. General purpose real-time image mosaicing, *appeared in the poster session of ICVSS*.

Bay H., Tuytelaars T., Gool L.V., 2006. SURF: Speeded Up Robust Features, *Proceedings of the ninth European Conference on Computer Vision*. 404-417.

Besl P., Mckay N.,1992. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, 239–256.

Brown M., and Lowe D. G., 2002. Invariant features from interest point groups. *British Machine Vision Conference, BMVC* 2002, Cardiff, Wales. 656-665.

Govender, N. 2009. Evaluation of feature detection algorithms for structure from motion. *3rd Robotics and Mechatronics Symposium (ROBMECH). Pretoria, South Africa*, 8–10,4.

Kang P., Ma H., 2011. An Automatic Airborne Image Mosaicing Method Based on the SIFT Feature Matching. *Multimedia Technology (ICMT)*, 155–159.

Ke, Y., Sukthankar, R., 2004. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. *Computer Vision and Pattern Recognition*, 506– 513.

Lazebnik, S., Schmid, C., and Ponce, J. 2004, Semi-Local Affine Parts for Object Recognition, *Proceedings of the British Machine Vision Conference*, 779–788.

Liu Z. An J., Jing Y., 2012. A Simple and Robust Feature Point Matching Algorithm Based on Restricted Spatial Order Constraints for Aerial Image Registration. *Geoscience and Remote Sensing, IEEE Transactions* V (50) Issue: 2, 514–527.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, 60, 2, 91–110.

Luo B., Hancock E. R., 2001. Structural graph matching using the EM algorithm and singular value decomposition, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, 1120–1136.

Martin A. Fischler, R., Bolles, C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Comm. Of the ACM*, vol. 24, 381–395.

Medioni G., Cohen I., Bremond F., Hongeng S., and Nevatia R., 2001. Event detection and analysis from video streams, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 873-889.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10, 27, 1615—1630.

Nakagawasai T., Saji H., 2011. Method of registering a time sequence of aerial images and a digital map using a satellite image. *IEEE Geoscience and Remote Sensing Society, Vancouver*, 3358–3361.

Sanroma G., Alquezar R., Serratos F., 2010. A Discrete Labelling Approach to Attributed Graph Matching Using SIFT Features. *20th International Conference on Pattern Recognition*. 954-957.

Schmid C., 1992. *Appariement d'Images par Invariants Locaux de Niveaux de Gris.* Doctoral Thesis, Institut National Polytechnique de Grenoble.

Sungho, K., Yoon K. J., Kweon I. S. 2006, Object Recognition Using a Generalized Robust Invariant Feature and Gestalt's Law of Proximity and Similarity, *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. 193.

Szeliski R., 2006. Image Alignment and Stitching: A Tutorial, *Handbook of Mathematical Models in Computer Vision*, Springer, 273–292.

Viola, P., Wells, W. M., 1997. Alignment by maximization of mutual information. *International Journal of Computer Vision*, *24* , 137–154.

Vivet M., Martinez B., Binefa X., 2011. DLIG: Direct Local Indirect Global alignment for Video Mosaicing. *IEEE Transactions on CSVT*. 21(12): 1869–1878.

Wei W., Jun H., and Yiping T., 2008. Image Matching for Geomorphic Measurement Based on SIFT and RANSAC Methods. *in Proc. CSSE (2)*.317-320.

Wyawahare, M. V., Patil, P. M., Abhyankar, H. K., 2009. Image Registration Techniques: An overview. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2, 1–5.

Xiong, Z. and Zhang, Y., 2009a. Image registration,In: Encyclopedia of Geography. *Sage Publication*.

Xiong, Z. and Zhang Y., 2010. A critical review of image registration methods, *International Journal of Image and Data Fusion*, 1:2, 137-158.

Zitova, B. Flusser, J., 2003. Image registration methods: a survey. *Image and Vision Computing*, 21, 977–1000.

APPENDIX

Figure 5: Some image mosaicing results.

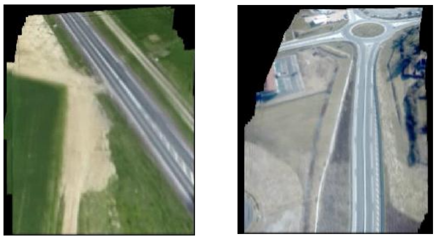
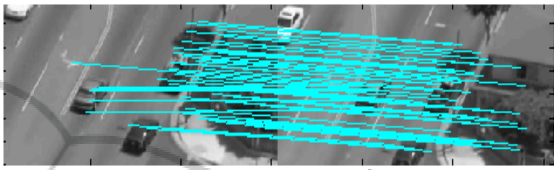
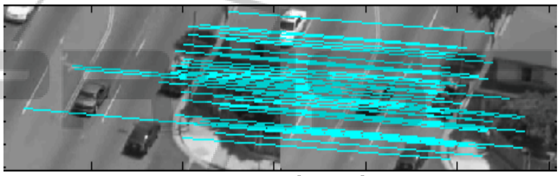


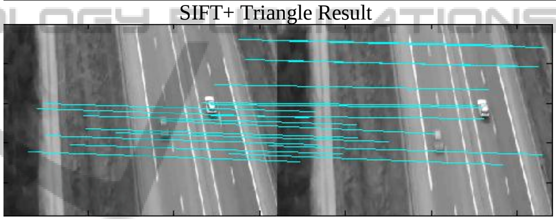
Figure 5: Some image mosaicing results (cont.).



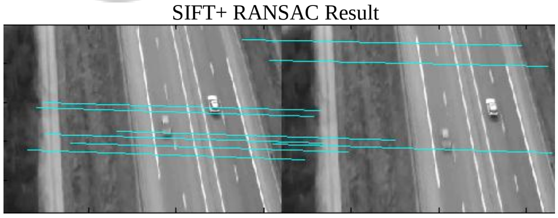
SIFT+ RANSAC Result



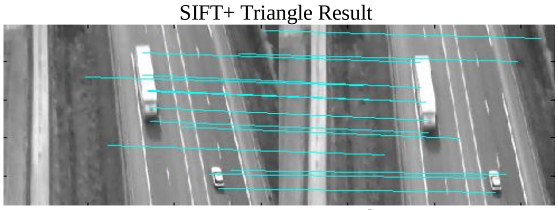
SIFT+ Triangle Result



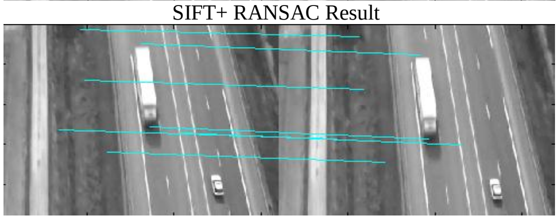
SIFT+ RANSAC Result



SIFT+ Triangle Result



SIFT+ RANSAC Result



SIFT+ Triangle Result

Figure 6: Results of the matching process.