Quantitative analysis of real-time image mosaicing algorithms

NEMRA Abdelkrim[1], *KHELLAL Atmane* [2], *SAYAH Houari* [3]
[1,2] *Ecole Militaire Polytechnique, B.P. 17 Bordj El Bahri, Algeria*
[3] *Service Informatique de l'armée, Algiers, Algeria*
karim_nemra@yahoo.fr

***Abstract* - Image registration and photo-mosaicing of related imagery from unmanned aerial vehicles (UAVs) is an active research topic. Aerial photography and mosaicing became crucial for surveillance and reconnaissance. It consists in aligning multiple images to construct a single large image of a 3D scene allowing the operator to view images that offers a wider field of view than standard images. Offline registration and mosaicing of collected images from UAVs has proven to give great results but is numerically intensive and somewhat slow. In this paper, a new approach for real time mosaicing is proposed based on new registration approach which reduces accumulated error and distortion of the current image. A quantitative analysis of the new approach is performed. This analysis allows the comparison of two images mosaic using some parameters to evaluate the quality, the Distortion rate and the speed of mosaicing.**

Keywords - Homography; Image mosaicing; KLT; SURF

I. INTRODUCTION

Recently, aerial photography has become an exciting new field of research. The limit of flight heights and the need of large photo cards have forced imaging experts to construct mosaic images from photographs that overlap to get a wider field of view. The construction of a large photo by image mosaicing consists in aligning of multiple images to form a single large picture of a 3D scene [1]. In our work, we are interested in photography and aerial surveillance, allowing the operator to view larges images with standard cameras [2].

Majumdar et.al.[9] presented a method for offline registration and mosaicing of collected images from UAVs. This method has proven to give great results but is numerically intensive and somewhat slow.

Instead, we use a new approach of image registration to construct a mosaic image from aerial photos. The proposed approach uses a new way of projection that keeps the current image as reference and the previous image mosaic will be registrated which reduces the accumulated error. The proposed approach surmounts the accumulated errors and reduces distortion in current image. Furthermore, a quantitative analysis is performed in this paper in order to evaluate the performances of image mosaic algorithm. This analysis allows an automatic comparison of two image mosaic algorithms; rather than visual comparison (qualitative). However, in this paper, we present a quantitative evaluation or a comparison using some parameters that tell us about the quality, the rate of Distortion and the speed of mosaicing.

II. PRINCIPLE OF IMAGE MOSAICING ALGORITHM

To build a mosaic image from a sequence of images (obtained from a set of images, a movie or a real video stream) three basic steps are required: image acquisition, image registration and perspective warping. These steps are shown in the following chart:

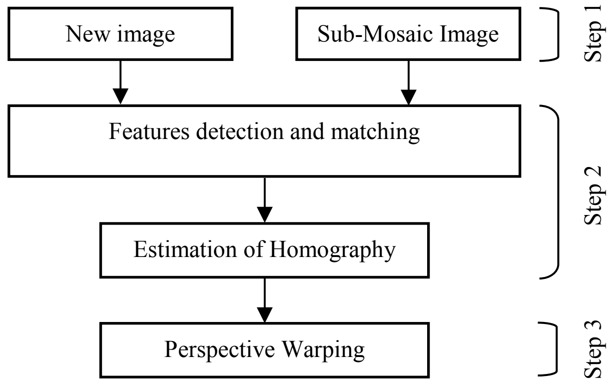

Figure 1. Image mosaicing algorithm.

A. Step 1: Image acquisition

In this step, two images are required: the first one is the new captured image noted I_k, and the second one is the Sub-Mosaic image noted I_s constructed with a set of previous images. A trivial choice of I_s is the entire image mosaic constructed by the $k-1$ first images. However, in this case the mosaic algorithm becomes computationally heavy especially when I_s become very large. As solution for this problem is to choose I_s as a rectangular region from the previous image mosaic with the same size as I_k.

B. Step 2: Image registration

The goal of this step is to estimate the geometric transformation for each acquired image frame following a reference image frame. To do so, invariant keypoints are detected and associated in both images Fig.3. Then, associated keypoints will be used to estimate the geometric transformation between frames. Acquired images can be affected by many kinds of 2D planar transformations that are shown in the following figure [3].

978-1-5386-6979-2/18/$31.00 ©2018 IEEE

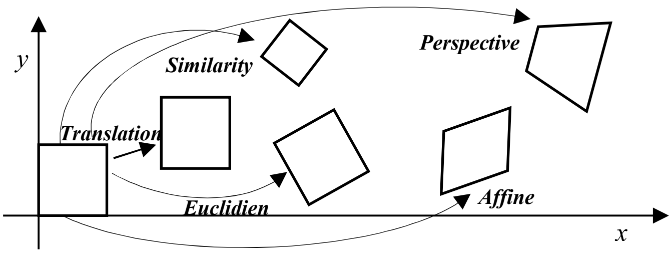

Figure 2. Basic 2D transformations.

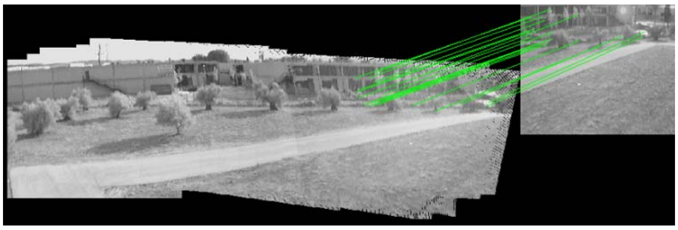

Figure 3. Features detection and association (sub-mosaic image and current image)

We note that the perspective transformation, called homography (noted *H*), is the most comprehensive of 2D planar transformation since it takes into account all the planar transformations (rotation, translation, scaling…). It is characterized by 8 DOF (Degrees Of Freedom). So both image points x' and x (projection of the same 3D point) in homogeneous coordinates are linked by:

$$x' \cong H \cdot x \tag{1}$$

Where $\cong$ means equal to a scale factor and H is a 3×3 matrix generally normalized by setting $h_{22} = 1$.

$$x' = \frac{h_{00}x \ \ +h_{01}y \ \ +h_{02}}{h_{20}x \ \ +h_{21}y \ \ +h_{22}} \quad y' = \frac{h_{10}x \ \ +h_{11}y \ \ +h_{12}}{h_{20}x \ \ +h_{21}y \ \ +h_{22}} \tag{2}$$

Note that each associated keypoints provides two equations (Eq.2). So to estimate the homography (08 unknown), we need at least four associated points. In our work, the homography matrix is estimated based on robust approach:

- Features detection and matching *:*
  In this step, we have implemented two approaches:
  1) Approach based on optical flow: Detect points of interest in an image using Harris detector and track them in the next image by estimating the optical flow using KLT algorithm (Kanade-Lucas-Tomasi tracker) [6].
  2) Approach based on local invariants: points of interest are first detected in the two images by SURF detector (Speeded-Up Robust Features). The detected points are then, matched is done by comparing SURF descriptors [7].

- Four pairs of point to point correspondences are at least required.
- Homography matrix estimation using the DLT algorithm (The Direct Linear Transformation algorithm) [4].
- Outliers rejection using RANSAC algorithm (RANdom SAmple Consensus algorithm) [5].
- Homography matrix estimated by RANSAC is used to initialize a non-linear estimator (the Levenberg Marquardt algorithm)[4], to minimize the back-projection error given by:

$$\sum_i \left(x_i' - \frac{h_{00}x_i+h_{01}y_i+h_{02}}{h_{20}x_i+h_{21}y_i+h_{22}}\right)^2 + \left(x_i' - \frac{h_{01}x_i+h_{11}y_i+h_{12}}{h_{20}x_i+h_{21}y_i+h_{22}}\right)^2 \tag{3}$$

B. Step 3: Perspective warping

Perspective wrapping consist in the application of projective transformation on the entire image (each pixel). For example, if the first frame is considered as a reference frame ($I_{ref} = I_0$). So to reset an acquired image frame at time $\boldsymbol{k}$ (I_k) we should use the homography matrix $H_{ref\to k}$ linking the images I_k and I_{ref}. Using the following formula:

$$H_{(ref\to k)} = \prod_{n=ref}^{n=k-1} H_{(n\to n+1)} \tag{4}$$

Then, the inverse transformation $\left(H_{(ref\to k)}\right)^{-1} = H_{(k\to ref)}$ is applied to the image I_k. This procedure is illustrated in Fig.4:

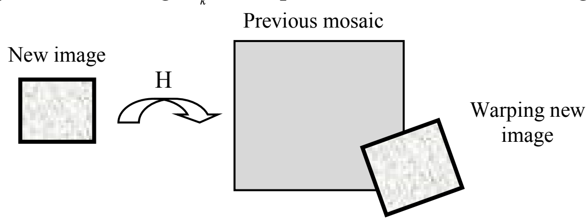

Figure 4. Warping new image technique.

Figure 5. Example using the warping new image technique.

Fig.5 shows an example of image mosaic where the first image is considered as image reference and the new image is wrapped. This projection technique suffers from cumulative error generated by the multiplication of intermediate homographies (Eq.4). In addition, the current image sometimes appears much distorted in the overall image mosaic (Fig.5). A simple alternative to overcome this problem is to warp the previous image mosaic [8]. To reset the current image I_k, simply we need the homography linking the current image I_k and the Sub-mosaic I_s. In this case I_s was chosen as the image I_{k-1} after wrapping. So we only need to estimate $H_{k-1\to k}$ which decreases significantly the accumulated error. Fig.6 shows the principle of the new mosaic strategy. Fig.7 shows an example of image mosaic constructed using this principle, as can be

978-1-5386-6979-2/18/$31.00 ©2018 IEEE

seen the current image (red square) doesn't sustain any transformation.

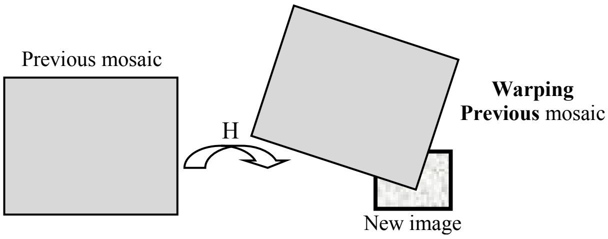

Figure 6. Warping previous mosaic technique.

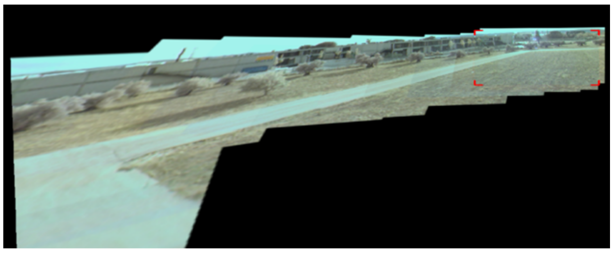

Figure 7. Example using the warping previous mosaic technique.

III. REAL TIME ASPECT IN IMAGE MOSAICING

Generally, cameras can operate at frequencies between 25 and 30 fps (frames per second). For 25 fps camera, image mosaicing algorithm (feature extraction and matching, image wrapping and image registration) should be run in less than 40 ms. Such a goal can never be achieved if SURF is used as a detector. Although SURF is described by its speed compared to SIFT (Scale-Invariant Feature Transform), it is still not suitable for real time application especially with large textured images. The computational time required to extract SURF parameters depends on the size and type of the acquired image, it depends also on its nature (textured images require more time than low textured images). In general, to extract ten points in an image (320 × 240) SURF can take about 60 ms (Intel CORE^{TM} i^3 with 4Go RAM Computer). To circumvent this time constraint, fast detector such as the KLT (Kanad Lucas Tomasi) tracking, can be used. However, despite its speed and simplicity it is very sensitive to noise, which leads to a large number of outliers. This latter affects significantly the estimation of homography matrix, therefore a low quality image mosaic will be obtained.

Thus, how can we construct a high quality mosaic image as fast as possible? Before answering this question we should answer the following question: should we really use all the acquired images by the camera? The answer is no, because a necessary condition to construct an image mosaic is the existence of an overlap between two successive images that contain at least four pairs of correspondences. In this case the required time between two frames which ensures a minimum overlap will increase. With this manner, sufficient time will be allowed to SURF to build the image mosaic.

IV. IMAGE MOSAIC ANALYSIS AND EVALUATION

In this section, we will define some parameters used in the analysis of two algorithms of image mosaic based on SURF detector/descriptor and KLT tracker.

A. *The Matching Score parameter*

To evaluate and compare different algorithms of mosaicing, we introduced a parameter called MS (Matching Score) which is defined as a ratio between the number of inliers (not rejected by RANSAC) and the total number of points used to estimate the homography. This is justified by the fact that a good mosaic is obtained by a good estimate of the homography.

$$MS = \frac{NI}{NI + NO} \tag{5}$$

Where: NI is the number of inliers and NO is the number of outliers. We can also define the MS_{moy} as an average of MS in a sequence of images.

$$MS_{moy} = (MS_1 + MS_2 + MS_3 + \cdots + MS_n)/N$$

Figure 8. Different MS values (left $MS = 1\%$, middle $MS = 56\%$, right $MS_{moy} = 89\%$).

Fig.8 shows an example of three images mosaic result. The Matching Score is evaluated for each image mosaic. As can be seen high values of MS reflect the good quality of the mosaic. After many experiences with different images, we found that for $MS \geq 50\%$ the mosaic image has good quality. In the general case the MS threshold depends on the value of RANSAC Threshold (RT), as shown in Fig.9, which illustrates the evolution of MS_{moy} according to the different RANSAC threshold. For $RT = 1.25$ we observe that good mosaic is obtained as soon as the MS exceeds 50 %.

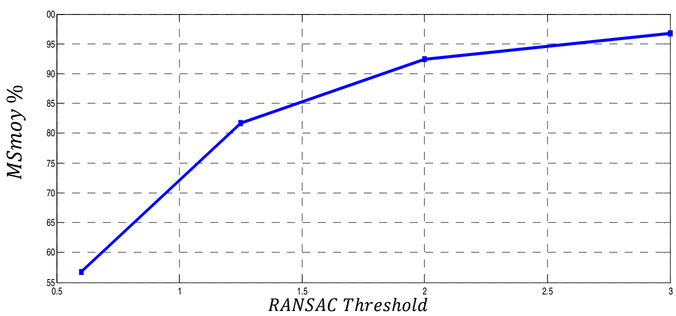

Figure 9. MS_{moy} according to RT

B. *The relative Distortion (RD)*

The relative Distortion of image mosaic is defined as a ratio between the number images where the MS is less than 50% and the total number of images. This parameter expressed in % provides information on the quality of the mosaic image.

C. *Number of Skipped Frame (NSF)*

The Number of Skipped Frame is the new sampling step of the video input. In other word, the frames which will participate in the construction of image mosaic are:

$$\{frame(1), frame(2), \ldots, frame(N)\}$$

With

$frame(i+1) = frame(i) + NSF$, $NSF \in \mathbb{N}^*$. Then the number of image used in mosaic algorithm $= \frac{total\ number\ of\ images}{NSF}$, when $NSF = 1$ this means the entire acquired images will participate in the construction of the image mosaic.

V. RESULTS AND ANALYSIS

The algorithms above mentioned are implemented (with Visual C ++ using computer vision library Open CV) and validated on several real and synthetic images. As well as video sequences and real video streams. Some results will be presented in this section.

A. *Real time constraint in image mosaicing*

Fig.10 summarizes Real time constraint in image mosaicing. The three dashed straight lines correspond to the ramps of a video stream at different speeds (25 fps, 20 fps and 15 fps). For cons, the curves in green and red correspond to the average execution time of mosaicing algorithm based SURF and KLT respectively. For example, if you want to ensure the constraint of real-time using mosaicing based SURF from a 25 fps video, choose an image jump (Number of Skipped Frames (NSF) or Frame Step (FS)) greater than or equal to 11 frames (corresponds to the intersection of the SURF curve (green) with the ramp 25 fps) and if we use KLT, choose number of skipped frames greater than or equal to 3 (corresponding to the intersection of the red curve with ramp 25fps). So intersection of the curve with the ramp correspond to the minimum number of skipped images (NSF_{min}) to meet the constraints of real-time.

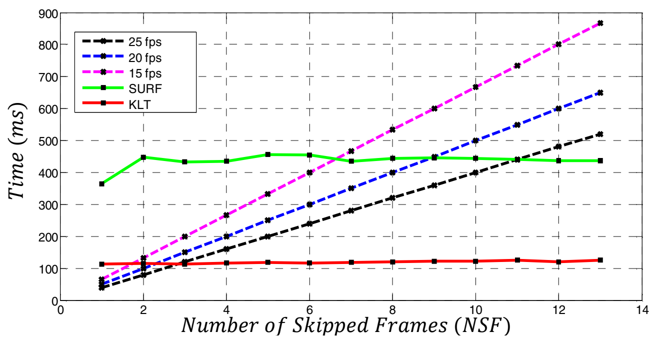

Figure 10. Real time constraints

As can be seen from Fig.10 But increasing the number of skipped frames has a great influence on the quality of obtained mosaic images. This is the purpose of the following paragraphs.

B. *Effect of NSF on MS_{moy}*

From Fig.11 we can observe that the MS_{moy} for SURF algorithm is 70%. It remains fixed, although the number of skipped frames (NSF) increases. This result can be explained by the robustness of SURF algorithm face significant geometric change. In this case, a sufficient condition to get good values of MS_{moy} is the presence of overlapping region that contains at least 4 inliers between two consecutive images. In the other hand, a necessary condition for KLT algorithm to track properly features is the slight changes between consecutives image. In other word KLT algorithm perform much better with small values of NSF , when this latter increases the quality of mosaic image with KLT decreases (Fig.11).

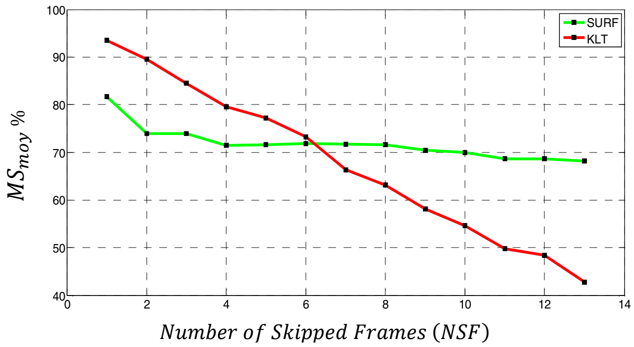

Figure 11. Effect of $\boldsymbol{NSF}$ on $\boldsymbol{MS_{moy}}$

C. *Effect of NFS on NI*

From Fig.12, we first observe that the number of inliers (NI) obtained with SURF algorithm is higher than KLT. We can observe also for both algorithm, that the number of inliers (NI) decreases when the number of skipped frames (NSF) increases. In other word, when the overlap between two successive images decreases.

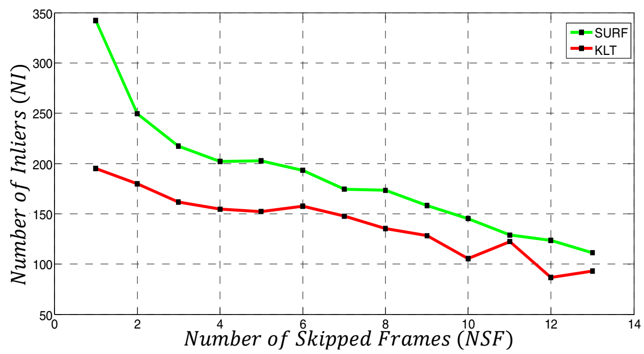

Figure 12. Effect of SF on Number of good matching.

D. Effect of NSF on RD

This parameter provides information on the quality of the mosaic image. As can be seen from Fig.13, when the NSF increases the relative distortion (RD) of image mosaic remains almost nil which confirm the robustness of the mosaic algorithm based on SURF. In the other hand, the distortion of the mosaic image constructed using KLT algorithm increases when the NSF increases. This can be explained by the sensitivity of KLT tracker to the time interval between frame s. For image mosaic based SURF; when $NSF > 10$ the overlap between consecutive images decreases, which increases the RD of the mosaic image.

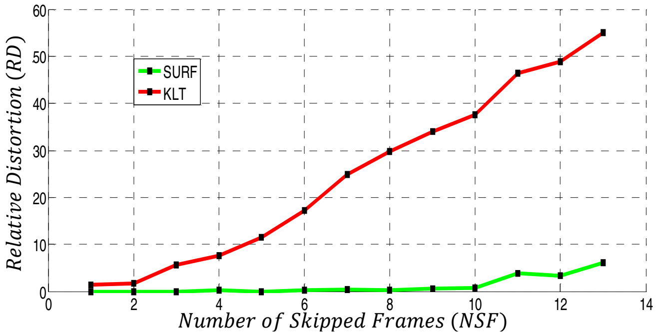

Figure 13. Influence of FS on RD

Fig.14 and 15 show an example of two results of images mosaicing in indoor and outdoor environment using SURF detector/descriptor.

Figure14. Image mosaicing in indoor environment

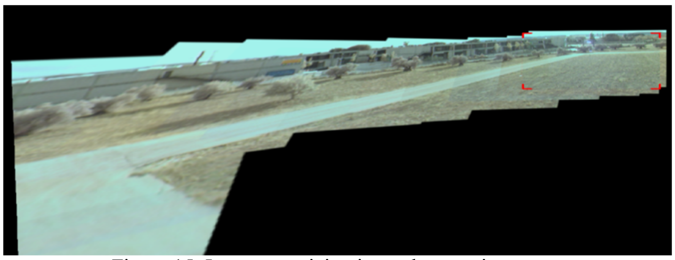

Figure 15. Image mosaicing in outdoor environment

VI. TEMPOREL ANALYSIS

The table below shows a temporal comparison between the two mosaic algorithms based SURF and KLT for two frames. As can be seen, mosaic based KLT algorithm is three times faster than mosaic based SURF algorithm. However, the total time for the entire mosaic algorithm based on SURF can be reduced by increasing the number of skipped frames (NSF). For example for$NFS = 3$, the number of images treated by mosaic based SURF algorithm is three time less than the number of images treated by mosaic based KLT algorithm. In this case the computational time of both algorithms is similar.

TABLE 1 COMPUTATION TIME COMPARISON

STEPS \ TIME	SURF		KLT	
	Time (ms)	*Time %*	*Time (ms)*	*Time %*
Image acquisition	1.92	0.53	1.46	1.34
Features extraction	73.49	20.24	4.81	4.43
Features association	188.51	51.94	2.06	1.90
Homography estimation	3.78	1.04	5.43	4.00
Image warping	95.22	26.23	94.96	87.32
Mosaic algorithm (*Mosaic time*)	362.93	100	108.75	100

VII. CONCLUSION

In this paper a quantitative analysis of image mosaicing algorithms is established. First, a new projection approach for image registration is applied. Then, two algorithms of images mosaicing are implemented and compared using accurate quantitative analysis. Experimental validation shows that the mosaicing algorithm based SURF gives much better results compared to KLT. Such a conclusion is expected since the robustness of SURF detector and descriptor.

VIII. REFERENCES

[1] Capel, D.P., "Image Mosaicing and Super-resolution", 2001.

[2] Heikkila, M. and M. Pietikainen, "An Image Mosaicing Module for Wide-Area Surveillance", 2005.

[3] Szeliski, R, "Image Alignment and Stitching", 2006.

[4] Harteley, R. and A. Zisserman, "Multiple View Geometry in Computer Vision", C.U. PRESS, Editor 2003.

[5] Fischler, M.A. and R.C. Bolles, "Random Sample Consensus A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", 1981.

[6] Rebière, N, Flavie, M, "Image Mosaicing using Local Optical Flow Registration", 2008.

[7] H. Bay, A.E., T. Tuytelaars, L. VanGool, "Speeded-Up Robust Features (SURF)". In ECCV, 2006.

[8] Indelman, V., P. Gur, and E. Rivlin, "Real-Time Mosaic-Aided Aerial Navigation : I. Motion Estimation". 2009.

[9] Majumdar, J., Vinay, S., Selvi, S.; Registration and Mosaicing for Images Obtained from UAV, IEEE International Conference on Signal Processing and Communications (SPCOM), 2004.