

Establishing a Common Coordinate View in Multiple Moving Aerial Cameras

Yaser Sheikh^a, Alexei Gritai^a, Imran Junejo^a, Robert Muise^b, Abhijit Mahalanobis^b, and Mubarak Shah^a

^aSchool of Computer Science, University of Central Florida, FL 32816, Orlando, USA

^bLockheed Martin Missiles & Fire Control, 5600 Sand Lake Road, FL 32819, Orlando, USA

ABSTRACT

A camera mounted on an aerial vehicle provides an excellent means of monitoring large areas of a scene. Utilizing several such cameras on different aerial vehicles allows further flexibility, in terms of increased visual scope and in the pursuit of multiple targets. The underlying concept of such co-operative sensing is to use inter-camera relationships to give global context to ‘locally’ obtained information at each camera. It is desirable, therefore, that the data collected at each camera and the inter-camera relationship discerned by the system be presented in a coherent visualization. Since the cameras are mounted on UAVs, large swaths of areas may be traversed in a short period of time, coherent visualization is indispensable for applications like surveillance and reconnaissance. While most visualization approaches have hitherto focused on data from a single camera at a time, as a consequence of tracking objects across cameras, we show that widely separated mosaics can be aligned, both in space and color, for concurrent visualization. Results are shown on a number of real sequences, validating our qualitative models.

Keywords: Co-operative Surveillance, UAV, Computer Vision

1. INTRODUCTION

The concept of a cooperative multi-sensor surveillance system, informally a ‘forest’ of sensors,¹² has recently received increasing attention from the research community. The idea is of great practical relevance, since cameras typically have limited fields of view, but are now available cheaply. Instead of having a high-resolution camera that surveys a large area, far greater flexibility and scalability can be achieved by observing a scene ‘through many eyes’, using a multitude of lower-resolution COTS (commercial off-the-shelf) cameras. Several approaches with varying constraints have been proposed, high-lighting the wide applicability of the concept. For instance, the problem of tracking across multiple *stationary* cameras with overlapping fields of view has been addressed by Lee *et al.*¹² and Khan *et al.*¹⁰ The assumption of overlapping fields of view was relaxed by Javed *et al.*⁸ and later by Rahimi *et al.*¹⁴ Motion, in terms of pan-tilt-zoom was first introduced to the ‘forest’ by Kang *et al.*,⁹ where correspondence was estimated between a combination of stationary and one moving camera (pan-tilt-zoom), using knowledge of the ground plane. In the proposed work, several objects are moving in a scene observed by several cameras, each mounted on (possibly) different UAVs, as shown in Figure 1. A compact representation of information is sought for the data obtained from these cameras, *without* any telemetry or calibration information. Obtaining calibration information in the current setup usually requires sophisticated equipment, such as a global positioning system (GPS) or an inertial navigation system (INS), perhaps with a geodetically aligned elevation map. Global positioning systems have a drawback, in that they can be blocked or intercepted, and the telemetry information provided by INS systems has a small but accumulating error involved. As a result, approaches based only on video data to recovering inter-frame relationships are particularly attractive as an alternative. Such a system finds obvious application in the monitoring of large areas where several aerial vehicles provide different views of the same scene. To achieve this we first estimate an optimal correspondence between object observed in the scene. However, since the cameras are moving and are often distant, direct appearance-based or proximity-based constraints cannot be used. Instead, we exploit constraints on the relationship between the motion of each object across cameras, proposing a novel trajectory similarity metric to test multiple correspondence hypotheses (without assuming any calibration information). Finally, while mosaics provide an excellent means of summarizing aerial video information from a *single* view, trying

Send correspondence to Yaser Sheikh

Yaser Sheikh: E-mail: yaser@cs.ucf.edu, Telephone: 1-407-823-4733

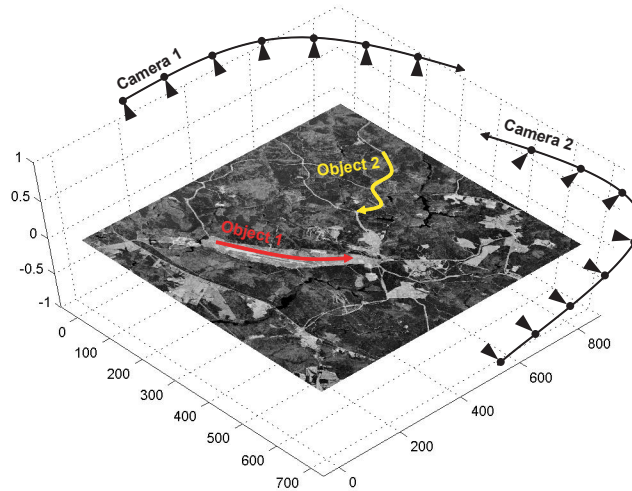


Figure 1. Multiple moving cameras observing a scene with moving objects. The context of this work is a scene where multiple moving objects are observed by multiple moving airborne cameras.

to simultaneously monitor information from several mosaics is awkward and inconvenient. Instead, we show that as a consequence of automatically tracking objects across multiple views, a *concurrent mosaic* can be computed summarizing the information from several aerial videos.

The notation used in this paper is as follows: there are N cameras, observing a scene with K objects. An object k present in the field of view of camera n is denoted as O_k^n . The imaged location of O_k^n at time t is $\mathbf{x}_{k,t}^n = (x_{k,t}^n, y_{k,t}^n, \lambda_{k,t}^n)^T \in \mathbb{P}^2$, the homogenous coordinates of the point* in sequence n . The trajectory of O_k^n is the set of points $\mathcal{X}_{k,\Delta t}^n = \{\mathbf{x}_{k,i}^n, \mathbf{x}_{k,i+1}^n, \dots, \mathbf{x}_{k,j}^n\}$, where Δt is the duration from frame i to frame j . For two cameras, a correspondence $c_{k,l}^{n,m}$ is an ordered pair (O_k^n, O_l^m) that represents the hypothesis that O_k^n and O_l^m are images of the same object. For more than two cameras, a correspondence $c_{i,j,k,\dots,l}^{m,n,o,\dots,p}$ is a hypothesis defined by the tuple $(O_i^m, O_j^n, O_k^o, \dots, O_l^p)$. Note that O_1^1 does not necessarily correspond to O_1^2 , the numbering of objects in each sequence is in the order of detection. Thus, the problem is to find the set of correspondences C such that $c_{i,j,k,\dots,l}^{m,n,o,\dots,p} \in C$ if and only if $O_i^m, O_j^n, O_k^o, \dots, O_l^p$ are images of the same object in the world. The rest of the paper is organized as follows: Section 2 describes the procedure to evaluate the similarity of object trajectories across the multiple moving views. In Section 3, the correspondence problem is posed and solved in a graph-theoretic framework. This is followed by the generation of the concurrent mosaic in Section 4. Results on controlled and real sequences are shown in Section 5, with conclusions in Section 6.

2. OBJECT TRACKING ACROSS MOVING CAMERAS

In this section, a means of evaluating the similarity of two trajectories observed in two different views is described. The principal assumption upon which this evaluation is constructed is that due to the altitude of the aerial camera, the scene can be well approximated by a plane in 3-space and as a result a homography exists between any two frames of any sequence. Within each sequence, object trajectories are obtained by first estimating the homography between consecutive frames, using a robust version of the direct approximation proposed by Mann and Picard.¹³ In order to detect independently moving objects, a cumulative frame-frame differencing approach was employed, as described in Irani *et al.*⁷ Finally, to correspond detected points across frames of a single sequence, we used the point correspondence approach proposed by Shafiq and Shah,¹⁵ a method allowing for occlusions, missed detections, and noise.

2.1. Measuring Similarity

Assuming all objects are visible in all moving cameras[†], after detection and subsequent tracking in each camera, there are K unlabeled trajectories in each of the N cameras, $\{\mathbf{x}_{k,t}^n\}$ for $t = 1$ to T , the duration of the tracks. We wish to correspond

*The abstraction of each object is as a point corresponding to the object centroid.

†This is not a required assumption, it is made here only for simplicity of description.

trajectories (and therefore objects) in each moving camera. Once motion has been compensated for in each sequence, the assumption of planarity dictates that a homography \mathbf{H} must exist between any two tracks that correspond. This constraint can be exploited to compute correspondence between trajectories. One straightforward approach to exploit this constraint is by finding the candidate match which minimizes the geometric distance or the transfer error. The transfer error for the set of correspondences is

$$\sum_t d(\mathbf{x}_{k,t}^n, \mathbf{H}\mathbf{x}_{l,t}^m),$$

where $d(\cdot)$ is some distance metric, like the Sum of Squared Difference or more robust metrics. Correspondence can be established by selecting the trajectory that minimizes this transfer error. However, this can involve a non-linear minimization in computing \mathbf{H} for each candidate match. Instead, in this work, the DLT (Direct Linear Transform) algorithm is used, which minimizes the norm $\|\mathbf{A}\mathbf{h}\|$ (see Proposition 1 for the definition of \mathbf{A} and \mathbf{h}). Since we are interested only in evaluating the similarity between two trajectories, and not necessarily in computing the homography between them, this method is appropriate.

Proposition 1 For an object in the world, the matrix \mathbf{A} constructed between the Trajectory $\mathcal{X}_{k,\Delta t}^p$ and the corresponding Trajectory $\mathcal{X}_{l,\Delta t}^q$, has a rank of at most 8, where

$$\mathbf{A} = \begin{bmatrix} \mathbf{0}^\top & -\lambda_{l,1}^q \mathbf{x}_{k,1}^p{}^\top & y_{l,1}^q \mathbf{x}_{k,1}^p{}^\top \\ \lambda_{l,1}^q \mathbf{x}_{k,1}^p{}^\top & \mathbf{0}^\top & -x_{l,1}^q \mathbf{x}_{k,1}^p{}^\top \\ -y_{l,1}^q \mathbf{x}_{k,1}^p{}^\top & x_{l,1}^q \mathbf{x}_{k,1}^p{}^\top & \mathbf{0}^\top \\ & \vdots & \\ \mathbf{0}^\top & -\lambda_{l,n}^q \mathbf{x}_{k,n}^p{}^\top & y_{l,n}^q \mathbf{x}_{k,n}^p{}^\top \\ \lambda_{l,n}^q \mathbf{x}_{k,n}^p{}^\top & \mathbf{0}^\top & -x_{l,n}^q \mathbf{x}_{k,n}^p{}^\top \\ -y_{l,n}^q \mathbf{x}_{k,n}^p{}^\top & x_{l,n}^q \mathbf{x}_{k,n}^p{}^\top & \mathbf{0}^\top \end{bmatrix}.$$

Proof Assuming that the condition of scene planarity holds, $s\mathbf{x}_{k,i}^p = \mathbf{H}_k^{p \rightarrow q} \mathbf{x}_{l,i}^q$, where $\mathbf{H}_k^{p \rightarrow q}$ is the homography between the trajectory of O_k^p and the trajectory of O_l^q . Thus, $\mathbf{x}_{k,i}^p \times \mathbf{H}_l^{p \rightarrow q} \mathbf{x}_{l,i}^q = \mathbf{0}$ which can be rewritten as $\mathbf{A}\mathbf{h} = \mathbf{0}$ where \mathbf{h} is \mathbf{H} in row-major form (for further details on the DLT see Zisserman *et al*⁵). Thus, since a null-space exists for \mathbf{A} , its rank can be at most 8.

Since the assumption of planarity does not hold exactly, and the trajectories are expected to contain some noise, the ninth singular value of \mathbf{A} will never be exactly zero. However, as a result of Proposition 1, the similarity between two trajectories can be measured by observing the condition number of \mathbf{A} , i.e. the ratio of the largest singular value to the smallest singular value of \mathbf{A} . Thus, the similarity between two trajectories can be computed using,

$$s(\mathcal{X}_{k,\Delta t}^p, \mathcal{X}_{l,\Delta t}^q) = \frac{\sigma_1^{\mathbf{A}}}{\sigma_9^{\mathbf{A}}} \quad (1)$$

where $\sigma_n^{\mathbf{A}}$ is the n -th largest singular value of \mathbf{A} .

2.2. Global Assignment of Labels

The problem of establishing correspondence between trajectories can be modeled within a graph theoretic framework. Consider first, the simplified case of several objects observed by two moving cameras. This can be modeled by constructing a complete bi-partite graph $G = (U, V, E)$ in which the vertices $U = \{u(\mathcal{X}_{1,\Delta t}^p), u(\mathcal{X}_{2,\Delta t}^p) \dots u(\mathcal{X}_{k,\Delta t}^p)\}$ represents the trajectories in Sequence p , and $V = \{v(\mathcal{X}_{1,\Delta t}^q), v(\mathcal{X}_{2,\Delta t}^q) \dots v(\mathcal{X}_{k,\Delta t}^q)\}$ represent the trajectories in Sequence q , and E represents the set of edges between any pair of trajectories from U and V . The completeness of the bi-partite graph belies the fact that any two trajectories may match hypothetically. The weight of each edge is the similarity of Trajectory $\mathcal{X}_{l,\Delta t}^q$ and Trajectory $\mathcal{X}_{k,\Delta t}^p$, as defined in Equation 1. By finding the maximum matching of G , we find a unique set of correspondence C' according to the following criterion,

$$C' = \arg \max_{C \in \mathcal{C}} \sum_{c_{l,q}^{k,p} \in C} s(\mathcal{X}_{k,\Delta t}^p, \mathcal{X}_{l,\Delta t}^q). \quad (2)$$

where \mathcal{C} is the solution space. Several algorithms exist for the efficient maximum matching of a bi-partite graph, for instance¹¹ or⁶ which are $O(n^3)$ and $O(n^{2.5})$ respectively.

Since the cameras are moving, the fields of view do not overlap all the time. For the correspondence of objects to be meaningful, the objects must be observed in both cameras simultaneously for some short duration. The minimum number of observations required to discern correspondence for two objects is four observations, i.e. both objects are observed simultaneously in the field of view for four (not necessarily consecutive) frames. Of course, since motion of cameras is smooth, the duration of overlap is usually significantly greater than four and this allows more stable computation of correspondence. Thus the base case in the online tracking assumes that some δ simultaneous observations have been observed (where $\delta \geq 4$). Then, as long as the number of objects remains constant, the similarity measure of Equation 1 between two hypothesized correspondences can be computed. With respect to Proposition 1, we have,

Corollary 1.1 All homographies mapping pairs of corresponding tracks in sequences p and q are equal, and are, in turn, the same homography that maps the mosaic constructed of Sequence p to the mosaic of Sequence q .

Corollary 1.1 provides us with a termination criteria. Since a characteristic of the correct correspondence is that all homographies between each pair of corresponded trajectories are equal, a final assignment of correspondence between two sequences is made based on the similarity metric of *all* objects,

$$s([\mathcal{X}_{m,\Delta t}^p \mathcal{X}_{n,\Delta t}^p \dots \mathcal{X}_{o,\Delta t}^p], [\mathcal{X}_{i,\Delta t}^q \mathcal{X}_{j,\Delta t}^q \dots \mathcal{X}_{k,\Delta t}^q]) \quad (3)$$

where $c_{i,j,\dots,k}^{m,n,\dots,o}$ is the correspondence found by the matching algorithm. By using all the points simultaneously to evaluate the ‘goodness’ of the solution, the relative position and orientation of each trajectory enforces a strong spatial constraint on correspondence.

3. CONCURRENT VISUALIZATION

The purpose of aerial surveillance is to obtain an understanding of what occurs in an area of interest. While it is well known that video mosaics can be used to compactly represent a single aerial video sequence, they cannot compactly represent several such sequences *simultaneously*. If, on the other hand, the homographies between each of the mosaics (corresponding to each aerial sequence) are known, a *concurrent* mosaic can be created of all the sequences simultaneously. Since each sequence is aligned to a single coordinate frame during the construction of individual mosaics, Proposition 1 provides us with the means to register mosaics from multiple sequences onto one concurrent mosaic. To this end, the known point-wise correspondences from object tracking can be used to compute the ‘inter-sequence’ homography (the null-vector of the matrix \mathbf{A} in Equation 3). Final alignment is then refined using direct registration.

3.1. Blending the Concurrent Mosaic

Although the moving cameras observe the same scene, the color values of corresponding points in the scene differ across the cameras. Figure 2 (a) shows a concurrent mosaic generated from two different sequences. Clearly, directly using these mosaics to create a concurrent mosaic causes noticeable artifacts, as shown in Figure 2. Assuming a Lambertian scene with a distant light source (the sun), the scene radiance, $L(\mathbf{X})$ depends only on material properties and the surface normal, i.e $L(\mathbf{X}) = \rho(\mathbf{X}) \mathbf{I} \cdot \mathbf{n}$, where $\rho(\mathbf{X})$ is the surface albedo at the world point \mathbf{X} , \mathbf{I} is the scene irradiance, and \mathbf{n} is the surface normal. Clearly, under the Lambertian model, the scene radiance does not vary with respect to the viewing direction. The image irradiance $I(\mathbf{x})$, in turn is linear in the scene radiance,

$$I_i(\mathbf{x}) = P_i e_i L(\mathbf{X}), \quad (4)$$

where P_i is an optical factor of camera i , and e_i is the exposure. Finally, let $M_i(\mathbf{x})$ be the intensity measurements at the world point \mathbf{X} available from the images obtained from camera i . These values are related to the image irradiance by the

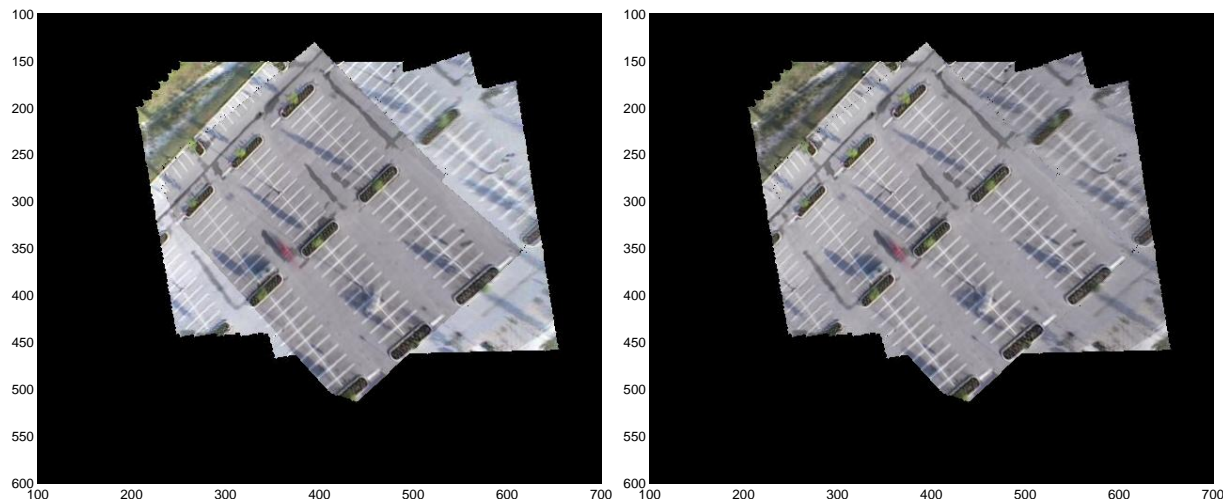


Figure 2. Concurrent visualization of two sequences. (a) Concurrent mosaic before blending, (b) Concurrent mosaic after blending.

radiometric response function f_i , $M_i(\mathbf{x}) = f_i(I_i(\mathbf{x}))$. Since f is a monotonically increasing function it is also invertible and we can define $g_i = f_i^{-1}$. We have

$$I_i(\mathbf{x}) = g_i(M_i(\mathbf{x})). \quad (5)$$

The source of intensity variation across images captured by different cameras can then only arise from the different radiometric response functions of the cameras. Grossberg *et al*³ proposed to relate the measured intensities in image captured by two cameras, $M_p(\mathbf{x})$ and $M_q(\mathbf{x})$ by a *intensity mapping function*, $M_p(\mathbf{x}) = G(M_q(\mathbf{x}))$. Since the scene radiance, L , does not vary with the viewpoint, from Equation 4,

$$\frac{I_p(\mathbf{x})}{P_p e_p} = \frac{I_q(\mathbf{x})}{P_q e_q}.$$

Using Equation 5 we then have,

$$M_p(\mathbf{x}) = g_p^{-1} \left(\frac{P_p e_p}{P_q e_q} g_q(M_q(\mathbf{x})) \right) = g_p^{-1}(k g_q(M_q(\mathbf{x}))) = G(M_q(\mathbf{x})). \quad (6)$$

This discussion can be extended to color (spectral reflectance) as well. The response of the i th sensor (red, green or blue) of camera p is expressed as,

$$M_p^{(i)}(\mathbf{x}) = f_p^{(i)} \left(\int_{\Lambda} \sigma_p^{(i)}(\lambda) I(\mathbf{X}, \lambda) d\lambda \right) \quad (7)$$

where λ is the wavelength, Λ is the range of visible wavelengths, σ is the spectral sensitivity of the i th sensor, I is the image irradiance. This approach of using three response functions for each color channel separately has been used by Grossberg *et al*,⁴ Shafique *et al*¹⁶ and Chang *et al*.² The three spectral response functions produce three corresponding *color transference functions*, G_r , G_b and G_g . color transference functions for color images are the analogue of intensity mapping function for grayscale images. However, the analogue is not direct since it was shown by Buchsbaum *et al*¹ that for any spectrally broadband color signal, each channel sensor produces outputs that are correlated. This correlation stems mainly (but not exclusively) from the spectral overlap of $\sigma^{(r)}$, $\sigma^{(g)}$ and $\sigma^{(b)}$. Thus, in order to model the color transference functions between the two views, it is important that the correlations between the channels be considered. Principally, a color transference functions matrix ought to be defined, however, by ignoring dispersion effects, we model the system as a multiple input single output system using multiple regression. We approximate each color transference functions by a

cubic trivariate polynomial,

$$r' = G_r(r, g, b) = \sum_{i+j+k \leq 3} a_{i,j,k}^{(r)} r^i g^j b^k - a_{0,0,0}^{(r)} + \epsilon, \quad (8)$$

$$b' = G_b(r, g, b) = \sum_{i+j+k \leq 3} a_{i,j,k}^{(g)} r^i g^j b^k - a_{0,0,0}^{(g)} + \epsilon, \quad (9)$$

$$g' = G_g(r, g, b) = \sum_{i+j+k \leq 3} a_{i,j,k}^{(b)} r^i g^j b^k - a_{0,0,0}^{(b)} + \epsilon. \quad (10)$$

where $a_{i,j,k}$ are the coefficients of the polynomial, $\{r, g, b\}$ and $\{r', g', b'\}$ are the color values the two images, and $\epsilon \sim \mathcal{N}(0, \sigma)$ is the i.i.d. random error. The property that $G_i(0, 0, 0) = 0$ is ensured by ignoring $a_{0,0,0}^{(i)}$. The error term exists because the available measurements invariably contain capture noise. Notably, the cross-product terms in this formulation take the correlation of the rgb axes into account. The coefficient of the polynomial can be solved through an overconstrained linear system of equations, since the homography is known between each pair of mosaics, where every point-to-point correspondence provides an $[r \ g \ b]^\top \leftrightarrow [r' \ g' \ b']^\top$ constraint.

Figure 2(b) shows a concurrent mosaic blended using the computed color transference functions. Furthermore, there are a number of phenomenon that are not modeled in the proposed model of the color transference functions, e.g. specular objects, saturation, or quantization in space and color, nominal misalignment. While accurate results have been obtained using the ordinary least squares (OLS) approach, since OLS is known to be sensitive to outliers, it is advisable to use a robust approach in solving the linear system, such as the least median square approach or iteratively reweighted least squares (used for the results in this paper). Of course, various further simplifications can be made to the multiple regression, such as assuming lower order models, or reducing the degree of the polynomial.

4. RESULTS AND CONCLUSION

In these experiments, two unmanned aerial vehicles (UAVs) mounted with cameras viewed real scenes with moving cars, typically with a smaller duration of overlap than the controlled sequence. Two sequences were recorded with three objects in the scene. Since the motion of aerial vehicles is far less controlled than that of real sequences, the duration of time in which a certain object is seen in both cameras is smaller. We show that despite the challenge of smaller overlap, object can be successfully tracked across the moving cameras. For the first experiment, a small number of frames were used, observing the motion of three moving cars. All three objects were simultaneously visible in the field of view for the entire duration of observation. The individual tracks of each sequence, on a single registered coordinate are shown in Figure 3 (a) and (b). The result of correspondence is shown in Figure 3 (c). Using this correspondence, the concurrent mosaic of the scene was generated, shown in Figure 2. In the second experiment, a longer sequence was used. This time, object exited and entered the field of view, and all three objects were only briefly visible together in the field of view. The individual tracks of each sequence, on a single registered coordinate are shown in Figure 4 (a) and (b). The result of correspondence is shown in Figure 4 (c). Using this correspondence, the concurrent mosaic of the scene was generated, shown in Figure 5.

Using multiple UAVs for aerial reconnaissance is an idea of wide applicability. While several algorithms have been proposed for rearranging the positions of the UAVs based on some sensors like GPS or INS for optimal coverage, object correspondence across multiple UAVs presents an interesting option once the ‘control loop’ is closed, namely that of rearranging multiple sensors using image information and object correspondence. Instead of a cost function of maximum coverage, or maximum overlap between UAVs, more intelligent cost functions based on object positions, proximity or object importance can be autonomously used.

ACKNOWLEDGMENTS

The authors would like to thank Donald Harper for engineering the hardware used in this project.

REFERENCES

1. G. Buchsbaum and A. Gottschalk, *Trichromacy, Opponent Colours Coding and Optimum Colour Information Transmission in the Retina*, Proceedings of the Royal Society of London, 1983.

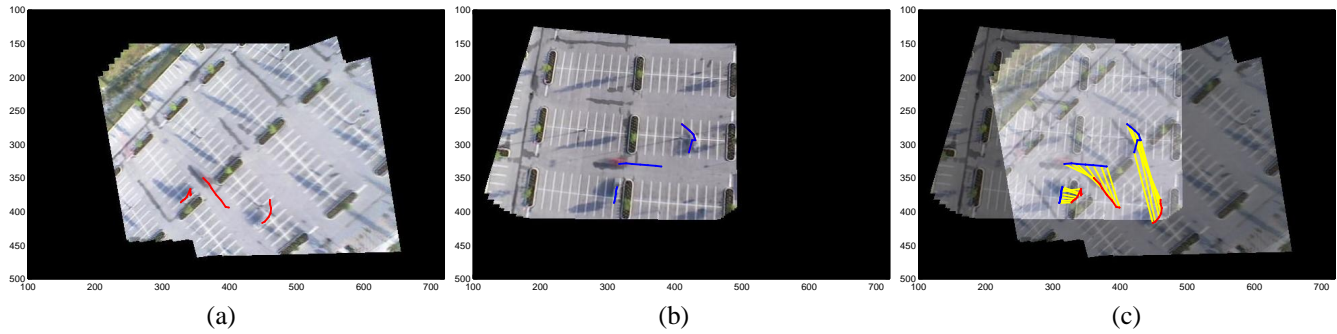


Figure 3. Object Correspondence between two sequences. (a) The red points show tracks of three objects detected and tracked in the first sequence (b) The blue points show the tracks of the same three objects detected and tracked in the second sequence and (c) Correspondences between the points are shown in a single plot by the yellow lines.

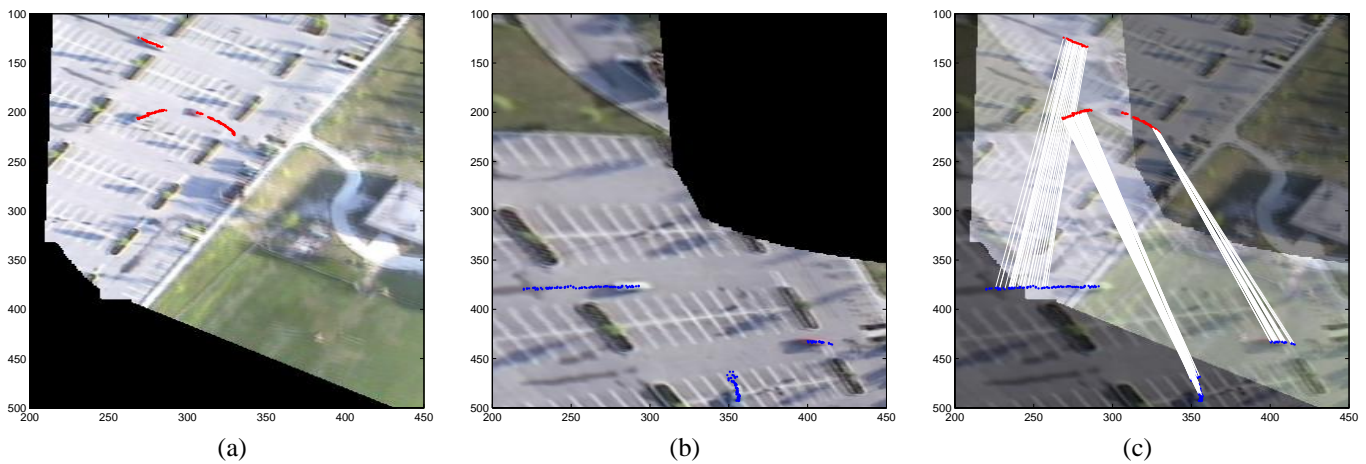


Figure 4. Object Correspondence between two sequences. (a) The red points show tracks of three objects detected and tracked in the first sequence (b) The blue points show the tracks of the same three objects detected and tracked in the second sequence and (c) Correspondences between the points are shown in a single plot by the white lines.

2. Y.-C. Chang, J. Reid, *RGB Calibration for Color Image Analysis in Machine Vision*, IEEE Transactions on Image Processing, 1996.
3. M. Grossberg and S. Nayar, *What can be Known about the Radiometric Response from Images?*, Proceedings of the European Conference on Computer Vision, 2002.
4. M. Grossberg and S. Nayar, *Modeling the Space of Camera Response Functions*, IEEE Transactions on Pattern Analysis and Machine Vision, 2004.
5. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.
6. J. Hopcroft and R. Karp, "A $n^{2.5}$ Algorithm for Maximum Matching in Bi-Partite Graphs", *SIAM Journal of Computing*, 1973.
7. M. Irani, B. Rousso, and S. Peleg, "Detecting and Tracking Multiple Moving Objects Using Temporal Integration", *Proc. European Conf. Computer Vision*, 1992.
8. O. Javed, Z. Rasheed, K. Shafiqe and M. Shah, "Tracking in Multiple Cameras with Disjoint Views", *IEEE International Conference on Computer Vision*, 2003.
9. J. Kang, I. Cohen, and G. Medioni, "Tracking Objects from Multiple Stationary and Moving Cameras", *Proceedings of the IEE Intelligent Distributed Surveillance System*, 2004.
10. S. Khan, M. Shah, "Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003.
11. H. Kuhn, "The Hungarian Method for Solving the Assignment Problem", *Naval Reserach Logistics Quarterly*, 1955.

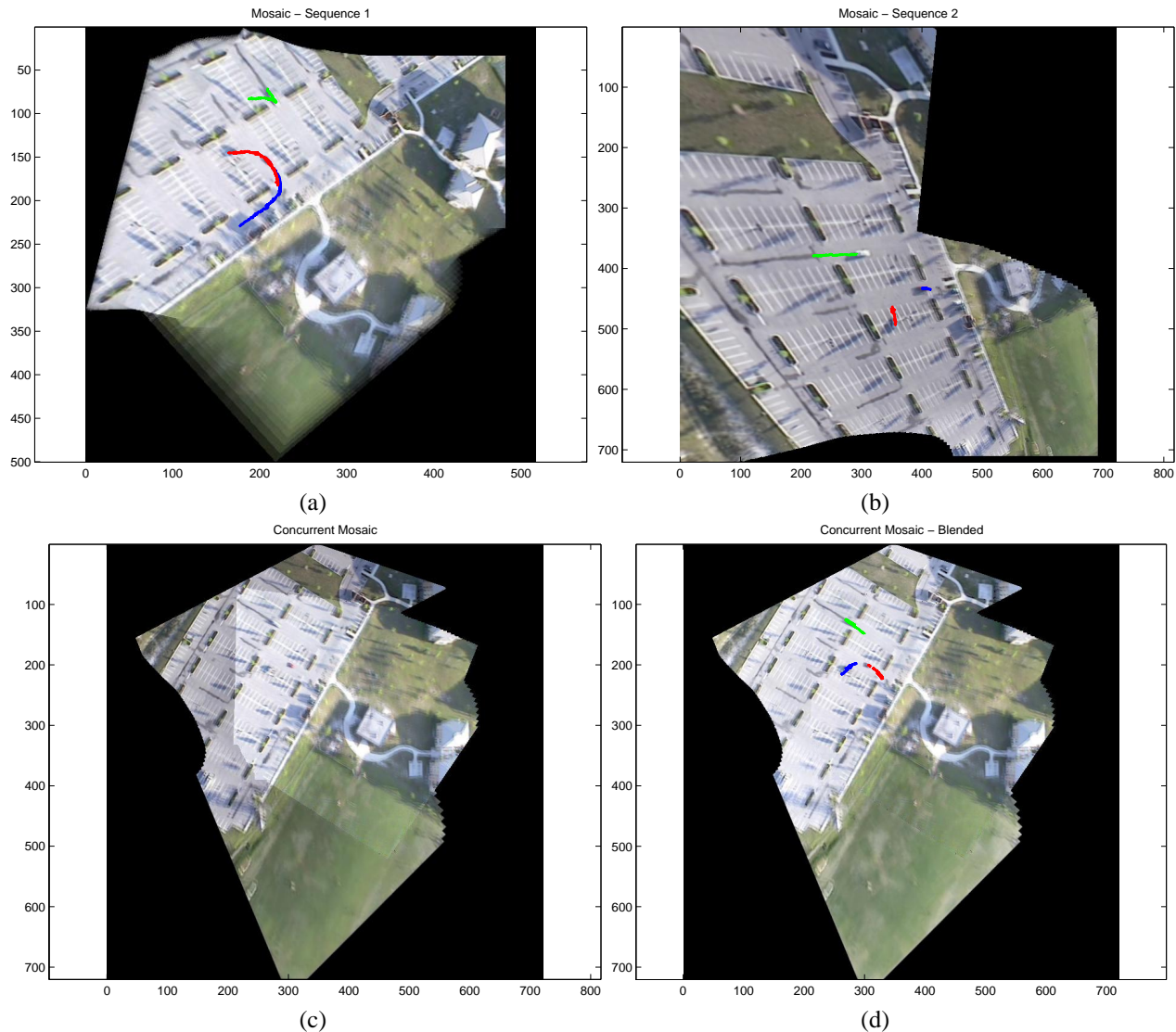


Figure 5. Concurrent visualization of two sequences. (a) Mosaic of Sequence 1 with object trajectories, (b) Mosaic of Sequence 2 with object trajectories, (c) Concurrent mosaic of both Sequence 1 and the warped mosaic of Sequence 2 (according to the homography computed from object correspondence). (d) Mosaics blended using a quadratic transfer function. Clearly, information about the objects and their motion is compactly summarized in the concurrent mosaic.

12. L. Lee, R. Romano, G. Stein, "Learning Patterns of Activity Using Real-Time Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.
13. S. Mann, R. Picard, "Video Orbits of the Projective Group: A Simple Approach to Featureless Estimation of Parameters", *IEEE Transactions on Image Processing*, 1997.
14. A. Rahimi, B. Dunagan, T. Darrell, "Simultaneous Calibration and Tracking with a Network of Non-Overlapping Sensors", *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.
15. K. Shafique and M. Shah, "A Noniterative Greedy Algorithm for Multiframe Point Correspondence", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.
16. K. Shafique and M. Shah, *Estimation of the Radiometric Response Functions of a Color Camera from Differently Illuminated Images*, IEEE International Conference on Image Processing, 2004.