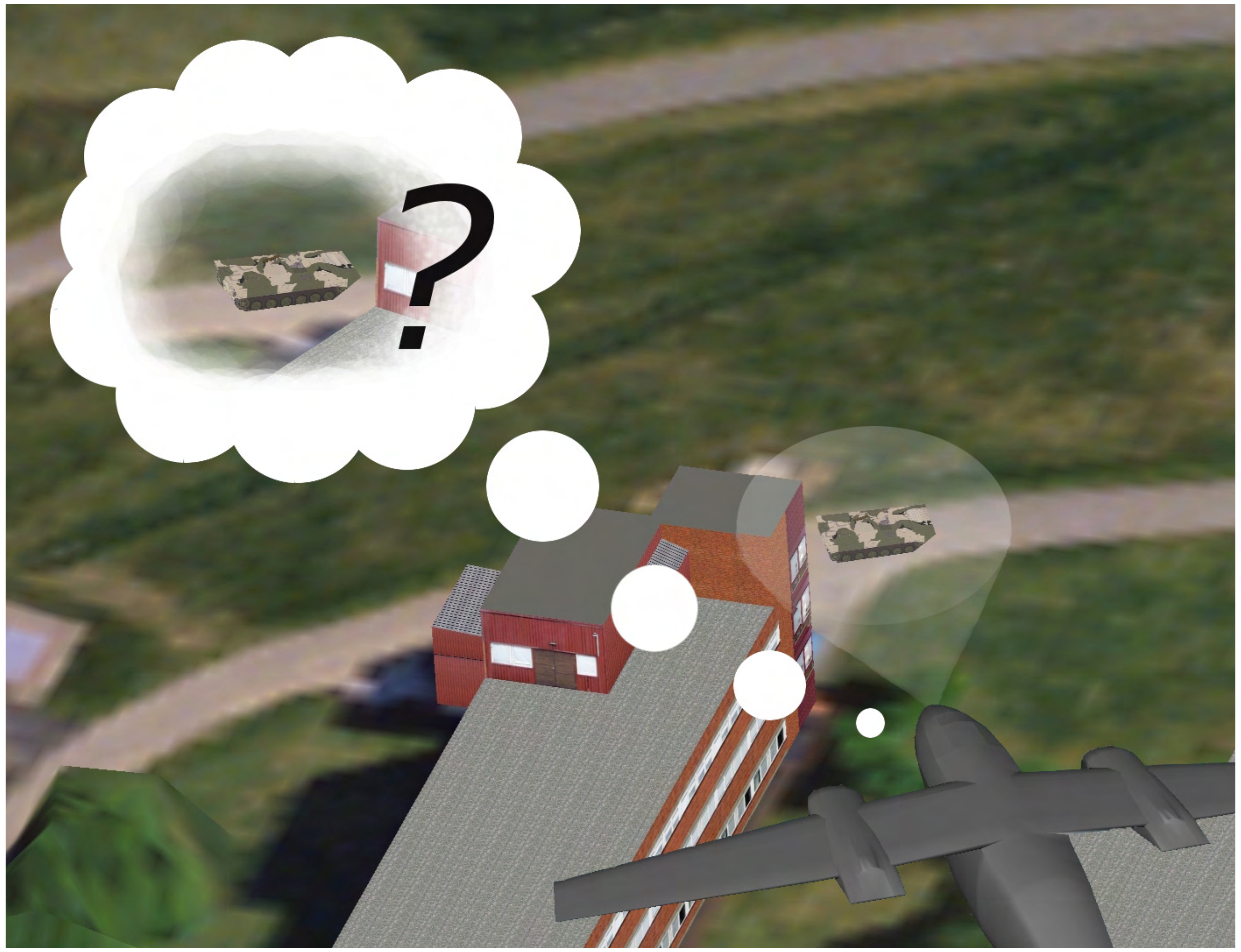
Re-identification of Previously Observed Targets in EO/IR-data from UAV

K-G STENBORG



FOI-R--2335--SE Technical report
ISSN 1650-1942 December 2007

Sensor Systems

FOI
Defence Research Agency
Sensor Systems
P.O. Box 1165
SE-581 11 Linköping

Phone: +46 13 37 80 00
Fax: +46 13 37 81 00

www.foi.se

K-G Stenborg

Re-identification of Previously Observed Targets in EO/IR-data from UAV

– A Survey

Rapportnr/Report no	FOI-R--2335--SE
Rapporttyp Report Type	Teknisk rapport Technical report
Utgivningsår/Year	2007
Antal sidor/Pages	21 p
ISSN	ISSN 1650-1942
Kund/Customer	FM
Forskningsområde Programme area	4. Sensorer och signaturanpassning 4. Sensors and Low Observables
Delområde Subcategory	42 Sensorer 42 Above surface Surveillance, Target acquisition and Reconnaissance
Projektnr/Project no	E3057
Godkänd av/Approved by	Jörgen Ahlberg

Totalförsvarets Forskningsinstitut FOI

Avdelningen för Sensorsystem

Box 1165

581 11 Linköping

Table of contents

Sammanfattning

I denna rapport har vi undersökt olika metoder för att återigenkänna fordon från en UAV. Återigenkänningsmetoden skall inte förlita sig på att fordonen finns i någon måldatabas innan uppdragets början. Tanken är att en UAV som följer ett målfordon skall kunna temporärt avbryta följningen för att utföra ett annat uppdrag och sedan återkomma, detektera fordon och bland dem återigenkänna målet som tidigare spårades. Därefter kan UAVn fortsätta med att följa målfordonet. Det är detta återigenkänningssteg i scenariot som har varit ämnet för denna rapport.

För att kunna återigenkänna målet måste vi samla särdrag för målet under följningsfasen. För att få återigenkänningen mer robust bör nya särdrag samlas in när de upptäcks eller uppstår. En bra återigenkänningsmetod bör kunna hantera förändringar i betraktningsvinkel, skala, ljusförhållanden och så vidare.

Vi presenterar några angripssätt för objektidentifiering och undersöker ett par återigenkänningsmetoder av fordon från UAV som är utvecklades inom VIVID-programmet.

I framtiden planerar vi att implementera en återigenkänningsalgoritm som använder en av VIVID-metoderna som grund men också införlivar andra identifieringssteg och metoder.

Nyckelord: återigenkänning, återidentifiering, identifiering, målföljning, UAV

Summary

In this report we have investigated different methods to re-identify vehicles from a UAV. The re-identification method should not rely on the fact that the vehicle is stored in a target database prior to the mission start. The idea is that a UAV that is tracking a target should be able to perform another task and then return, detect vehicles, and among them re-identify the previously tracked vehicle. Thereafter the UAV can continue with the tracking of the vehicle. It is the re-identification step in this scenario that has been the topic of this report.

To be able to re-identify the target we must collect target features during the continuous tracking. For a more robust re-identification new target features should be assembled when they appear. A good re-identification method should be able to handle changes in pose, scale, lighting conditions and so on.

We have presented some approaches for object recognition and investigated a couple of re-identification methods of vehicles from UAV that was developed in the VIVID program.

In the future we will implement a re-identification algorithm that uses one of the VIVID methods as a model but also incorporates other recognition steps and methods.

Keyword: re-identification, recognition, target tracking, UAV

1 Introduction

Automatic detection and tracking of moving ground vehicles can be done from an Unmanned Aerial Vehicle (UAV). Using EO/IR-sensors in a pan-tilt gimbal the UAV can track a target vehicle continuously. If the UAV receives a temporary assignment it might have to move the gimbal or temporary fly away so that the target vehicle is no longer within the field of view. After completing the new assignment the UAV will continue with the tracking of the target vehicle. To do so the target must be detected and re-identified. It is this re-identification that is the topic of this report.

1.1 Scenario

For our scenario no prior knowledge of the target vehicle is assumed even if the target signature can be contained in a database. We also assume that more than one vehicle is located within sensor range of the UAV.

The first steps in tracking a vehicle are:

1. Detection of a moving or stationary vehicle
2. Manual decision of which vehicle that is the target (target designation)
3. Collection of features from the target vehicle
4. Continuous tracking of the target vehicle and collecting of features/model refinement

During step 4 a new assignment is prioritised by the UAV. During this assignment the target vehicle is not within field of view. After some time the UAV once more priorities the target vehicle. Then the following steps are taken:

5. Predict an uncertainty area for the position of the target vehicle (gating)
6. Detect vehicles within the area
7. Re-identify the target among detected vehicles
8. Update the features of the target vehicle
9. Continuous tracking of the target vehicle

During the last step the UAV is again ready to carry out new assignments, see figure 1. Note that steps 3 and 4 are done simultaneously, as are steps 8 and 9.

In this report we will focus on steps 3 and 7 in the scenario, the collecting of target features and the re-identification of the target vehicle. This means that we assume that methods exist for detecting and tracking vehicles. In step 3 the features are collected and they can be used to build a model, stored in a database, or be processed to other forms that are adopted for the re-identification. For the re-identification in step 7 we will, from a set of detected vehicles, identify which of them, if any, that is the target vehicle.

The extent of time between when the continuous tracking is temporary stopped (step 4) and the target vehicle is once more prioritised (revisit time, step 5) can be from less than one second up to several minutes.

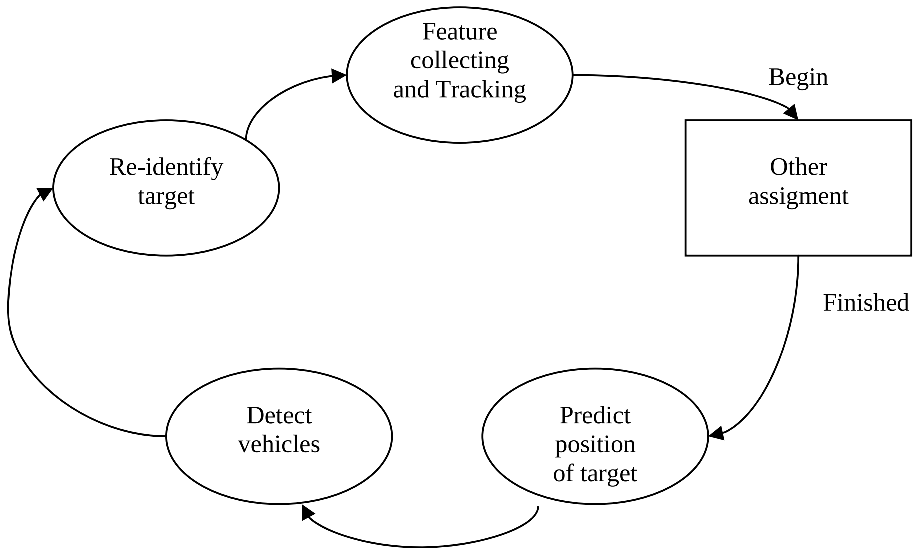


Figure 1: After the UAV has finished another assignment the position of the target is predicted and vehicles are detected. Thereafter the target is re-identified among the detected vehicles. Finally the target tracking and collection of new target features can be performed.

By using this re-identification the UAV can track more than one vehicle even when they are not all within the field of view. Multiple-target tracking (MTT) has been used for IRST (infra-red search and track) and radar for many years [Blackman]. One type of MTT is track-while-scan (TWS) that is a system that with regular intervals scan a predefined area with a sensor. TWS will both search for new targets and track targets that are already known.

The trade off between how much resource that should be spent on tracking vehicles and how often we should search for new vehicles is an interesting issue for TWS. If we track a target on a road with sparse traffic we can spend much time on the re-identification and tracking of the target and little time on the search for new vehicles. In an urban area we might instead need to allocate much more time to the search for new vehicles and will have less amount of time for the re-identification and tracking of targets. This way the re-identification will be less robust for areas with many vehicles.

1.2 Visual conditions

When the UAV is trying to re-identify the target vehicle, visual conditions may have changed since the UAV was last continuously tracking the target. Conditions that can be different between the detected target vehicle and the saved features are:

- Bearing, depending on position of the UAV compared to the vehicle
- The orientation of the vehicle
- Scale of the vehicle dependent on the distance change between the UAV and the vehicle
- Resolution of the vehicle on the sensor, also dependent on the distance between them
- Vehicle articulation

- The irradiance from the vehicle
- The color may appear different depending on the lighting conditions
- Shadows on and around the vehicle that may have changed
- Specular reflexes on the vehicle
- Operational state of the vehicle
- The vehicle can be partly occluded during the re-identification

Some of these conditions may be less obvious in IR compared to EO and for some it may be the opposite. For the IR sensor there is a memory embedded since the temperature of the vehicle will not change as fast as the conditions (lighting, shadows and specular reflexes) for the EO. We would like to use a re-identification algorithm that is robust to changes in all of these conditions.

The first four points (Geometrical rotation, scale and resolution) we refer to as the collection geometry. Usually all parts of the collection geometry can be dealt with at the same step in the matching algorithm.

If the vehicle is articulated the geometrical features of the vehicle will change. This means that either the geometrical features will not be a reliable for matching or that the matching must somehow be able to compensate for the non-rigid body.

The irradiance and spectral distribution of the vehicle will change for different lighting conditions. If these differences are not too large most matching algorithms should have no problem compensating for them.

Shadows may have changed because the vehicle is now in another orientation relative to the sun (or other light source). The shadows can also be stronger or weaker dependent on the amount of clouds that are covering the sun and the vehicle can (partly) be in the shadow of other structures. The sun (and other light sources) may also create specular reflections on the vehicle that will change the visual appearance of it.

Operational state will be of importance for IR sensors since the temperature of the vehicle will depend on whether the engine is on or off. If the vehicle is moving the friction will cause the wheels to be warmer compared to if the vehicle stands still. Other dynamic effects will in a similar way influence the temperature of the vehicle.

If the vehicle is partly occluded and this prevents matching, the UAV should hopefully be able to get a non occluded view of the vehicle later on and be able to make a correct matching.

1.3 Occlusion

During continuous tracking the target may be temporarily occluded, for example by buildings in an urban environment. This also calls for a form of re-identification when the target is again visible. For targets that are only occluded for short moments the location, orientation and other conditions will not have changed much and therefore the re-identification will for most cases be easy. Two ways of handling occlusion are joint probabilistic data associative filter (JPDAF) [Rasmussen] and robust statistics [Gross2006].

2 Recognition of 3D objects

In this section we will give a quick overview of some of the recognition approaches that are possible. We use the four different classes for recognizing query objects from training data described in [Forsyth].

Pose-consistent Approaches

Geometrical mechanisms are used to identify a sufficient number of matches between image and model features. For example, an alignment technique can be used with a tree search matching process that use the fact what given only a few matches the object pose are given. This will in turn limit the position of the features for further matches. One well known alignment approach is the one given in [Basri].

Another method is to use a small number of points to compute a feature vector that is independent of the viewpoint. The vector can be used to index a hash table storing all models.

Template Matchers

Images from different views, orientations and scales of an object are collected. This type of observed data is stored for all objects that we are interested in. A classifier is then used to determine whether a query object is any of the objects in the observed database. Not only geometrical features but also color, brightness and so on are used for the matching.

A problem with this type of method is that the background will cause errors since the background in the observed data will be compared to the background of the query object. Therefore a segmentation step must be included before the classifier (for both the training data and query).

Relational Matchers

If the object is articulated a probabilistic model can be used to describe the relation between the components of the object. This is, for example, a good recognition approach for humans, since they can have their arms and legs in different positions. The arms and legs can be recognized individually and then the model will see if they are placed in an appropriate configuration.

Aspect Graphs

The aspect graph gives a representation of from which viewpoints the appearance for an object will be qualitatively different. By calculating the aspect graphs for the training data and compare them with the measured aspect graph of the query object the similarity between them are given. Unfortunately, it has shown to be difficult to calculate exact aspect graphs in practice. This will affect the robustness of recognition methods based on the graphs.

Both the pose-consistent and template matchers are recognition approaches that can be used for our case. Relation matchers are of interest for articulated vehicles, but since most vehicles have a rigid body this type of matchers are not priorities. They might be of interest further on in the development. Finally, the aspect graph method seems to be interesting but since vehicles have similar shapes the aspects graphs might be very similar. An aspect graph method should not do the re-identification by itself but be a part of an algorithm containing more recognition methods.

Using more than one matching method will increase the accuracy for the re-identification. For example, a shape matching method can be combined with a color matching method for more robustness.

3 Related areas

There are applications other than target tracking from a UAV that are of interest for automatic re-identification. In this section we will describe some of them.

3.1 Traffic surveillance

If cameras placed along a roadway can identify passing vehicles, parameters such as travel time, origin, and destination can be observed [Sun]. For each camera all vehicles will be observed from the same angle. Also, the cameras are usually placed in such way that they all have similar viewpoint of the traffic. These two conditions make the re-identification less complicated than from a UAV since the expected view angle for each camera can be predicted. Since similar viewpoints of the vehicle already exist from previous camera observations, a matching will be more successful compared to the case that no similar view of the vehicle exists.

One simple yet efficient way for vehicle re-identification is of course license plate matching for the cases when that is possible. Of more interest to us is vehicle re-identification using multi-detector fusion [Sun].

3.2 Face recognition

In this large research field [Zhao] identification of faces are being investigated. Since humans are very sensitive to faces it is difficult to reach the same level of identification for an automatic system [Sinha]. The first systems dealt with faces which were nicely placed with the front to the camera and good lightning conditions. More resent research deals with faces that are not directly looking into the camera [Gross2004].

3.3 Other areas of re-identification

Other areas where re-identification is being used are for matching fingerprints and lumber identification in the different steps in sawmills.

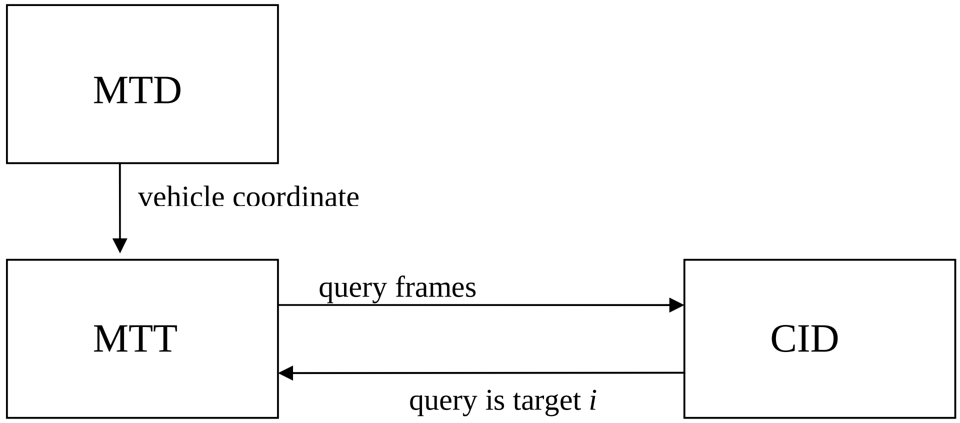


Figure 2: A vehicle is detected by MTD and MTT will transmit a few query images of it to the CID. The CID compares the similarity between the query and targets. The CID decision is then transmitted back to the MTT.

4 Re-identification methods developed in VIVID

In this section we will first give a short description of the VIVID program and then describe the two different re-identification algorithms (CID) that was proposed. Finally a third re-identification algorithm will be described.

4.1 VIVID

In the DARPA program Video Verification of Identity [Strat][Arambel] the three areas of Moving Target Detection (MTD), Moving Target Tracking (MTT) and Confirmatory Identification (CID) are addressed. VIVID is designed to be able to track at least 3 vehicles from a UAV at the same time even if they are widely separated. Both EO and IR sensors can be used.

The MTD is used to detect vehicles on the ground. Detected vehicles will be tracked by the MTT that also collects frames of the vehicles. The MTT then transmits the frames of the query vehicles to the CID to identify the targets among them, see figure 2.

The Confirmatory Identification (CID) is such an important component in VIVID that two different approaches have been developed. One of these is developed by SAIC Imagery Technology System Division and the other by Sarnoff Corporation.

4.2 SAIC Confirmatory Identification

The Science Applications International Corporation CID method [Guarino][Yue2005][Yue2006] separates the look of the object (**spatial mode**) and how the object behaves (**temporal mode**). The method contains two steps, Target Model Development and Target Verification.

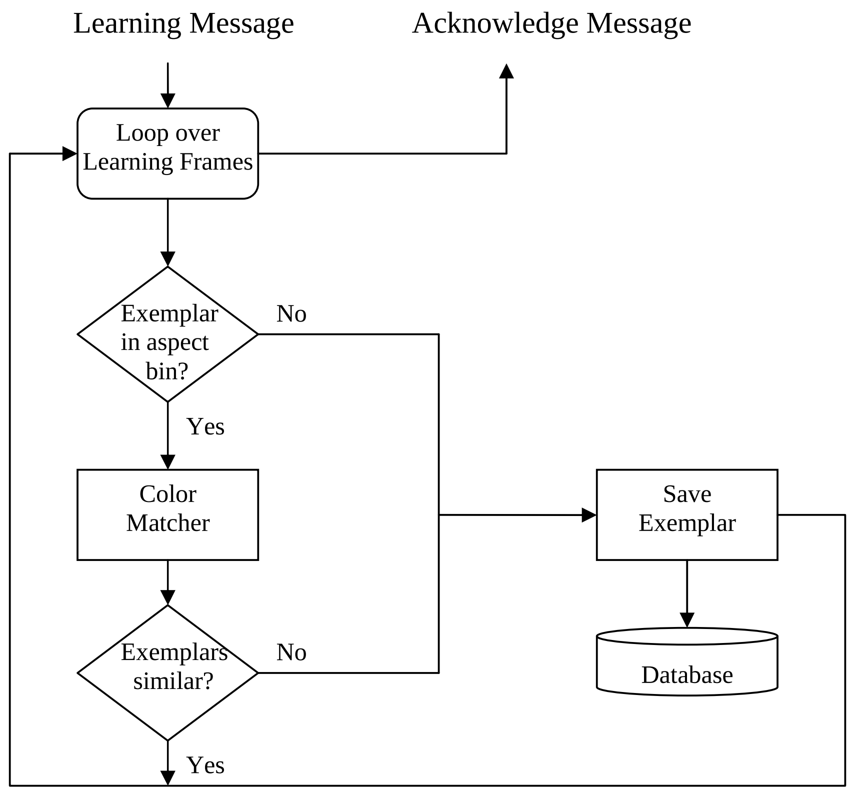


Figure 3: Frames in a learning message is transmitted to the CID. The learning frames are stored in a database if similar frames are not already saved.

4.2.1 Target Model Development

During the Target Model Development process the models for the CID are being built. This is done on-the-fly and the target only needs to be viewed for a relatively short time, about 1-3 seconds of video. The MTT sends the learning video sequence to CID along with metadata containing information about the target location in the frame, resolution, aspect angle and depression angle.

The CID use a sample selection to determine which of the frames that should be stored in the example database. The database is divided into 72 bins based upon aspect angle (5 degrees of aspect in each bin). The metadata information is used to determine the correct bin for every learning frame. If the corresponding bin in the database already contains a frame the similarities between that frame and the learning frame is determined using a Color Matcher (that will be described in next section), see figure 3. Since the Color Matcher is rotationally variant it will measure the difference both in color and the spatial features. Since frames from different views are stored in a database this is a type of template matcher.

4.2.2 Target Verification

The Target Verification is used to confirm the targets identification. The MTT sends a query sequence of less than one second of video to the CID. The MTT also transmits a list of possible targets. The CID then matches the query with the targets in the database and reports a likelihood ratio for each possible target and the likelihood that the query is none of them. The target verification is described in figure 4.

The color and spatial matching works best if the orientation of the query and target examples are similar. Therefore, the example frames that are selected for the match should have similar viewpoints. The algorithm does allow symmetry between left and right, but not front to back. Normalization is done by using histogram stretching on the luminance component. This gives a better distribution of values and a mean value that is closer to the middle value of the luminance value range. The normalization also gives better quantization for the color matching and more consistently defined edges for the spatial matching.

The color matching is based on color co-occurrence histogram [Chang]. To reduce the complexity of the algorithm the colors are quantized to a small number of levels. Three-dimensional Color co-occurrence matrices (CCMs), with dimensions of the number of colors, by the number of colors, by the number of distances (*nDist*), are calculated. The CCM can be either rotationally invariant or variant depending upon how they are calculated.

To calculate the matching score between two images the intersection of the diagonal (measure of uniform color areas) and non-diagonal (measure of image texture) CCM elements are determined separately. The query and example CCM are given by *qCCM* and *lCCM* will then give the intersection values

$$diag = \sum_{k}^{nDist} \sum_{i}^{n} \min(qCCM_{ii}^{k}, lCCM_{ii}^{k})$$

$$ndiag = \sum_{k}^{nDist} \sum_{i}^{n} \sum_{j}^{n} \min(qCCM_{ij}^{k}, lCCM_{ij}^{k}), i \neq j \ .$$

These two intersection values are then combined in a weighted sum that gives the matching score

$$d = 1 - \frac{(dWt * diag) + (ndWt * ndiag)}{nDist * (dwt + ndWt)} \ .$$

The weights *dWt* and *ndWt* can be used to control whether the diagonal or non-diagonal features should be prioritised. The matching score *d* are between 0 and 1 since the CCM are normalized. The lower matching score we get the better the similarity between the images is.

To match the spatial features of the query and target examples they need to have the same collection geometry and that is provided by a View Synthesis. The View Synthesis uses a homography to first calculate the needed transform and then use it to change the aspect angle, depression angle and resolution to create a similar view of a selected target example as the viewpoint of the query vehicle.

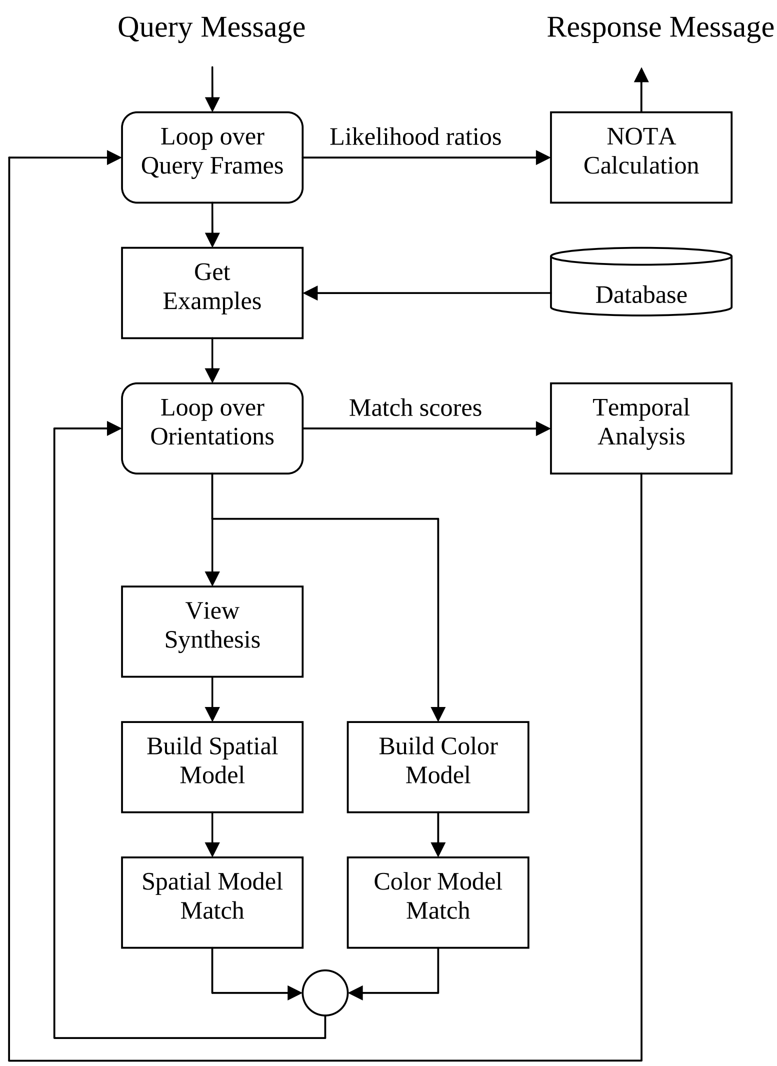


Figure 4: Query target is compared with learning exemplars.

For the spatial matching geometrical features (horizontal and vertical ridges) are extracted from the vehicles. The ridges are both located as lines on the vehicles and the outlines of the vehicles. The algorithm used for the spatial matching is based upon the Distance Transform. The Distance Transform matcher will match pose, object type, localized spatial information and multiple attribute types between the query vehicle and a vehicle in the database. The matching will also be insensitive to partial occlusion.

The Temporal Analysis combines spatial matching with the temporal behaviour of the target by using a generalization of a hidden Markov model called Dynamic Bayesian Net (DBN). The DBN has 72 states (representing 72 possible orientations) for each frame. Matches are not done at all 72 states because of the computational load. Instead, only matches for orientations closer than 20 degrees from the orientation of the query vehicle (given by the metadata) are performed. This will reduce the number of matches from 72 to 9.

The probability for a match at orientation *i* for frame *t* are given by

$$\alpha_t(i) = \sum_j \alpha_{t-1}(j) A_{j,i} e^{-S_{i,t}} \ .$$

$S_{i,t}$ is the combined color and spatial scores for orientation *i* at time *t* and $A_{j,i}$ is the temporal model that represent the probability that the target can move from orientation *j* to *i* during one time step. This will result in the likelihood

$$L_t = \left(\sum_i \alpha_t(i) \right)^{1/t}$$

at time *t*. The likelihood for each target in the database is reported to the MTT. After the MTT has confirmed that the target is the desired one the CID uses the verification observations to update the target spatial and temporal models.

The CID is also called by the MTT within intervals during the continuous tracking to ensure that the target is the correct one. This will also give the CID new and updated model databases that are more robust over time.

4.3 Sarnoff Confirmatory Identification

The CID method that Sarnoff developed [Guo2005a] [Guo2005b] [Guo2007] uses descriptors such as regions, lines and points to match the query target with the vehicles in the database. The region features are used to delineate the object and the line and point features provide accurate localization and robust alignment across disparate views.

The algorithm can be described by two steps - within-sequence object mask generation and across-sequence image alignment and matching.

4.3.1 Within-sequence object mask generation

An adoption of the robust blob features [Forssen] is used in combination with the Earth Mover’s Distance (EMD) [Rubner] to generate object masks.

First the blob features are used to extract homogeneous regions in the image. These regions can be overlapping and since the regions do not follow the exact shape of the object the extraction is robust to scale, appearance and view changes.

In the next step region matching will be performed by the EMD on the blob features from a set of images from a sequence. The EMD is based on a solution of the *transportation problem* and compares the similarity between two distributions. The name Earth Mover’s Distance comes from the fact that it can be used to calculate the least amount of work needed to fill some holes in the ground with the content of some heaped piles of earth. In a similar way the EMD can be used to measure the similarity between two images with respect to the color, location and area of the blob features in the images. The properties of the metric will measure two images as similar even if the number and size of the blobs are different between them as long as the sets of blobs describe similar features from a similar object. Iteratively small blobs in one image can be merged based on the homogeneity in the other image. After the iterative step the set of blobs in the two images should be similar with corresponding blobs if both images contain the same object. Blobs corresponding to the background should be removed if the background is different between the two images. After this step the EMD cost will indicate the similarity between the two objects.

The third step will create the object masks. This is done by using a similar technique as the region matching. The mask is created by the union of the blobs that are left after the iterative step. A dilation operation is applied to expand the mask by 20% so the whole object will be contained inside it. We can note that shadows cast by the vehicles are also contained in the blob union and therefore also inside the mask.

4.3.2 Across-sequence image alignment and matching

To deal with pose and appearance changes for the matching between the query and learning images corner-like and line features are used. Edges are detected and corresponding lines classified as being in the X-axis or Y-axis (with positive Y-axis as approximated driving direction and the Z-axis perpendicular to the ground). Lines that are overlapping or close by are combined to form longer lines. Band images are formed by collecting [Y R G B] profiles along a small band along each edge. This band images are used as descriptors. The normalized correlations between the band images are now calculated to give the line correspondence between a query and learning image (one correspondence for the X-lines and one for the Y-lines). A “quasi-rigid” alignment is finally calculated for the two images. This can be seen as a pose-consistent approach that at the same time makes a color matching because of the band images.

Using the “quasi-rigid” alignment a matching score between the two images is calculated. The score contains different parts such as geometrical structure, normalized color correlation and color similarity. The three matching score parts are combined using weights.

Finally, for the sequence-to-sequence matching key frames are chosen from the query and learning sequences. Each query key frame is matched with each learning key frame. Temporal constraints can be used for deciding on the best match between the sequences. Another possibility is to use the best key frame matching score.

4.4 Trifocal Tensor approach

The VIVID program has also resulted in one more re-identification method [Shao]. This method simultaneously tracks and performs recognition of the vehicles (i.e., does not have the MTT/CID-separation). The re-identification algorithm uses trifocal tensors [Avidan], a method for describing the relation between different views of a 3D object. To create a trifocal tensor three images with different perspectives of the object is needed.

A 3D transform algorithm is used during the tracking. For each frame the trifocal tensor is estimated using the previous frames. Since the tensor estimation is based on noisy measurements the initial trifocal tensor will not be good but the algorithm will improve the tensor for each new frame.

Two different approaches for the object recognition are used - one that is based on previously built trifocal tensors from training data and one that use texture mappings of the target vehicles. The images uses image warping to get the same pose before the similarity between the query and target model is measured. They show that the trifocal tensor recognition method is the one that converges fastest.

To be fully robust this re-identification method need to track and build the trifocal tensor of the query target for quite a long time. Compared to the few frames that is needed for the CID-methods in VIVID this is a problem for the trifocal tensor approach.

5 Summary, Conclusions and Future work

In this report we have investigated different methods to re-identify vehicles from a UAV. The idea is that a UAV that is tracking a target should be able to perform another task and then return, detect vehicles, and among them re-identify the previously tracked vehicle. Thereafter the UAV can continue with the tracking of the vehicle. It is the re-identification step in this scenario that has been the topic of this report. To be able to re-identify the target we must collect target features during the continuous tracking. For a more robust re-identification new target features should be assembled when they appear. A good re-identification method should be able to handle changes in pose, scale, lighting conditions and so on.

We have presented some approaches for object recognition and investigated a couple of re-identification methods of vehicles from UAV that was developed in the VIVID program.

In the future we will implement a re-identification algorithm that uses one of the VIVID methods as a model but also incorporates other recognition steps and methods.

References

[Arambel] Pablo Arambel, Matthew Antone, Constantino Rago, Herbert Landau and Thomas Strat, *A Multiple-Hypothesis Tracking of Multiple Ground Targets from Aerial Video with Dynamic Sensor Control*, Proceedings of the Seventh International Conference on Information Fusion, vol. 2, pp. 1080-1087, Jun 2004.

[Avidan] Shai Avidan and Amnon Shashua, *Novel View Synthesis by Cascading Trilinear Tensors*, IEEE Transactions on Visualization and Computer Graphics, vol. 4, no. 4, pp. 293-306, 1998.

[Basri] Ronen Basri, *Viewer-Centered Representations in Object Recognition: A Computational Approach*, In C.H. Chen, L.F. Pau, and P.S.P. Wang (Eds), Handbook of Pattern Recognition and Computer Vision. World Scientific Publishing Company, Singapore, chapter 5.4, pp. 863-882, 1993.

[Blackman] Samuel S. Blackman and Robert Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, ISBN 1-58053-006-0, Aug 1999.

[Chang] Peng Chang and John Krumm, *Object Recognition with Color Cooccurrence Histograms*, IEEE Conference on Computer Vision and Pattern Recognition, June 1999.

[Forssen] Per-Erik Forssén, *Low and Medium Level Vision using Channel Representations*, Dissertation No. 858, ISBN 91-7373-876-X, Linköping University, Sweden, March 2004.

[Forsyth] David A. Forsyth and Jean Ponce, *Computer Vision: A Modern Approach – International Edition*, Prentice Hall, ISBN 0-13-191193-7, 2003.

[Gross2004] R. Gross, S. Baker, I. Matthews and T. Kanade, *Face Recognition Across Pose and Illumination*, Handbook of Face Recognition, Stan Z. Li and Anil K. Jain, ed., Springer-Verlag, Jun 2004.

[Gross2006] Ralph Gross, Iain Matthews and Simon Baker, *Active Appearance Models with Occlusion*, Image and Vision Computing, vol. 24, no. 6, pp. 593-604, 2006.

[Guarino] David Guarino, Bradley Walls and Eric Miles, *Confirmatory Identification of Targets in Video – Final report*, Video Verification of Identity (VIVID): Automated Video Processing for Unmanned Aircraft - A compilation of scientific papers and technical reports that summarizes the accomplishments of the DARPA VIVID program, Phase 1, Edited T. Strat and L. Hollan, 2005.

[Guo2005a] Yanlin Guo, Steve Hsu, Harpreet S. Sawhney, Rakesh Kumar and Cen Rao, *Video Object Fingerprinting for Confirmatory Identification (CID) - Sarnoff's Contribution to Large Area Multi-Target Tracking (MTT) in Aerial Videos under Phase I of the DARPA VIVID Program*, Video Verification of Identity (VIVID): Automated Video Processing for Unmanned Aircraft - A compilation of scientific papers and technical reports that summarizes the accomplishments of the DARPA VIVID program, Phase 1, Edited T. Strat and L. Hollan, 2005.

[Guo2005b] Yanlin Guo, Steve Hsu, Ying Shan, Harpreet Sawhney and Rakesch Kumar, *Vehicle Fingerprinting for Reacquisition & Tracking in Videos*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR'05, vol. 2, pp. 751-768, June 2005.

[Guo2007] Yanlin Guo, Steve Hsu, Harpreet S. Sawhney, Rakesch Kumar and Ying Shan, *Robust Object Matching for Persistent Tracking with Heterogeneous Features*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 5, May 2007.

[Rasmussen] Christopher Rasmussen and Gregory D. Hager, *Probabilistic Data Association Methods for Tracking Complex Visual Objects*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 560-576, 2001.

[Rubner] Yossi Rubner, Carlo Tomasi and Leonidas J. Guibas, *The Earth Mover's Distance as a Metric for Image Retrieval*, International Journal Computer Vision, vol. 40, no. 2, pp. 99-121, 2000.

[Shao] Jie Shao, Shaohua Kevin Zhou and Rama Chellappa*, Appearance-Based Tracking and Recognition Using the 3D Trilinear Tensor*, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04), vol. 3, May 2003.

[Sinha] P. Sinha, B. Balas, Y. Ostrovsky and R. Russell, *Face Recognition by Humans: 19 Results All Computer Vision Researchers Should Know About*, Proceedings of the IEEE, vol. 94, no. 11, pp. 1948-1962, Nov 2006.

[Strat] Thomas M. Strat, Lois C. Holland and William E. Thomas, *Video Verification of Identity (VIVID)*, Video Verification of Identity (VIVID): Automated Video Processing for Unmanned Aircraft - A compilation of scientific papers and technical reports that summarizes the accomplishments of the DARPA VIVID program, Phase 1, Edited T. Strat and L. Hollan, 2005.

[Sun] Carlos C. Sun, Glenn S. Arr, Ravi P. Ramachandran and Stephen G. Ritchie, *Vehicle Reidentification using Multidetector Fusion*, IEEE Transactions on Intelligent Transportation Systems, vol. 5, no. 3, pp. 155-164, Sept 2004.

[Yue2005] Zhanfeng Yue, Antonio Trias and Rama Chellappa, *Synthesis of Novel Views of Moving Objects Airborne Video*, Video Verification of Identity (VIVID): Automated Video Processing for Unmanned Aircraft - A compilation of scientific papers and technical reports that summarizes the accomplishments of the DARPA VIVID program, Phase 1, Edited T. Strat and L. Hollan, 2005.

[Yue2006] Zhanfeng Yue, David Guarino and Rama Chellappa, *Moving Objects Verification in Airborne Video*, IEEE International Conference on Computer Vision Systems (ICVS), New York, NY, Jan 2006.

[Zhao] W. Zhao, R. Chellappa, A. Rosenfeld and P.J. Phillips, *Face Recognition: A Literature Survey*, ACM Computing Surveys, pp. 399-458, 2003.