Low Resolution Vehicle Re-Identification Based on Appearance Features for Wide Area Motion Imagery

Mickael Cormier¹, Lars Wilko Sommer¹,², and Michael Teutsch¹

¹Fraunhofer IOSB, Fraunhoferstrasse 1, 76131 Karlsruhe, Germany
²Vision and Fusion Lab, Karlsruhe Institute of Technology (KIT), Adenauerring 4, 76131 Karlsruhe, Germany

{mickael.cormier|lars.sommer|michael.teutsch}@iosb.fraunhofer.de

Abstract

The description of vehicle appearance in Wide Area Motion Imagery (WAMI) data is challenging due to low resolution and renunciation of color. However, appearance information can effectively support multiple object tracking or queries in a real-time vehicle database. In this paper, we present a systematic evaluation of existing appearance descriptors that are applicable to low resolution vehicle re-identification in WAMI data. The problem is formulated as a one-to-many re-identification problem in a closed-set, where a query vehicle has to be found in a list of candidates that is ranked w.r.t. their matching similarity. For our evaluation we use a subset of the WPAFB 2009 dataset. Most promising results are achieved by a combined descriptor of Local Binary Patterns (LBP) and Local Variance Measure (VAR) applied to local grid cells of the image. Our results can be used to improve appearance based multiple object tracking algorithms and real-time vehicle database search algorithms.

1. Introduction

Wide Area Motion Imagery (WAMI) sensors are a great source of data for wide area aerial surveillance systems. Thousands of moving vehicles on the ground can be tracked simultaneously due to the sensors’ large ground coverage of several square kilometers per frame. Various applications can be derived such as traffic monitoring, driver behavior analysis, or queries in a real-time vehicle database. However, there is one major drawback when analyzing individual vehicles in WAMI data: currently available datasets such as the WPAFB 2009 dataset [39] are monochromatic and each vehicle covers only about 10 x 20 pixels in the image. Together with the low frame rate of 1-2 Hz these are specific properties of WAMI data that are used to limit the large amount of data.

Hence, there is only little information available about the appearance of each individual vehicle. This is a problem since appearance based multiple object tracking algorithms or database queries applied to moving vehicles in WAMI data strongly suffer from ambiguities [30]. This is especially the case in urban areas with dense traffic [21, 8]. As a result, authors who consider vehicle appearance often use simple appearance features such as gray value histograms [26, 25] or blob features [34, 31] to support vehicle re-identification between two frames with sufficient robustness. Such vehicle re-identification methods are helpful to support the data association problem and the track linking problem in multiple object tracking [8, 25] and traditional re-identification tasks such as database queries [44, 19, 17].

In this paper, we focus on appearance based vehicle re-identification in WAMI data. The problem is formulated traditionally by finding a query vehicle in a list of candidates that is ranked w.r.t. their matching similarity [41]. We use the ground truth (GT) of the WPAFB 2009 dataset to generate a large set of queries for our experiments. Our main contribution lies in the systematic evaluation of different appearance descriptors that are applicable to WAMI data such as Local Binary Patterns (LBP) or the covariance descriptor as well as suitable distance measures. So far, no such evaluation exists in the literature. We expect that our results can help prospective authors to choose powerful appearance descriptors that improve multiple object tracking algorithms and real-time database search algorithms.

The remainder of this paper is organized as follows: literature related to WAMI and Person re-identification is reviewed in Section 2. The experimental data is described in Section 3 and the descriptors for appearance based re-identification in Section 4, respectively. Experimental results are given in Section 5. We conclude in Section 6.
2. Related Work

Recent literature on WAMI data processing mainly focuses on vehicle detection and tracking [10]. Since vehicle appearance is challenging to be considered for vehicle detection and tracking due to the low resolution [31], many authors simply ignore appearance information and use background suppression in combination with connected-component labeling to detect moving objects [29, 33, 43, 21, 8]. This method often leads to noisy detection results [37] and multiple object tracking is then performed using complex graph models in order to robustly create tracks despite of potentially suboptimal detections [34]. However, vehicle detection algorithms based on appearance features and machine learning do need context knowledge in order to achieve high robustness against false alarms [36, 22]. As a result, motion information is most important in WAMI tracking and appearance is used incidentally in order to improve data association, track linking, or persistent tracking [3, 27, 31, 8, 13]. The used descriptors are either gray value histograms [26, 25] or blob features [34, 31].

In general, solving the re-identification problem is crucial for many applications such as long-term multi-camera tracking [15] and forensic search [4]. In particular, in many surveillance applications a vehicle or a person disappearing from the camera’s field of view (FOV) is required to be matched in one or more other views from other cameras at different physical locations over a period of time, and be discriminated from several other candidates that are visually similar [18]. This problem has been widely formulated as a retrieval or recognition problem [9]. In this paper, the overall re-identification process is considered as a ranking problem [32, 40] under the assumption of a one-to-many matching problem in a closed-set. This means that only one candidate of the gallery set corresponds to the query. More realistic scenarios such as one-to-many matching problem in open set and many-to-many matching problem have been addressed recently [45, 11].

3. WAMI Vehicle Re-Identification Dataset

While there are several public datasets available for the field of person re-identification [9], no similar public datasets for the field of vehicle re-identification in WAMI data are available so far. This work is built upon the WPAFB 2009 dataset, a public WAMI dataset provided by the U.S. AFRL [39]. The WPAFB 2009 dataset consists of 3,072 images with an average size of 26,000 × 21,000 pixels that are captured at a frame rate of 1.2 Hz. The GT includes (but is not limited to) the position and a persistent identifier for each vehicle in each frame. Catrambone et al. [12] reported remarkable issues with the original GT such as position anomalies or slightly drifting tracks. They released a revised version of the GT that was not available at the time of the writing of this paper, unfortunately.

For our experiments, we use a subset of the WPAFB 2009 dataset and modify the original GT in order to evaluate appearance based vehicle re-identification. First, a challenging region of 2,278 × 2,278 pixels with dense traffic is cropped from the original images. This region is similar to AOI 40 that is used by Basharat et al. [8]. A pair of directly succeeding frames is selected for every tenth frame of the cropped region. This new set of 100 pairs of frames well represents the variability of the original dataset in terms of varying distance of the camera to the scene and changes in illumination or camera gain. There are 14,623 vehicles in total with 72 vehicles per pair of frames in average. Rectangular bounding boxes are placed around the vehicles and rotated based on their orientation provided by the motion direction of the GT track. Additionally, the bounding boxes are resized manually in order to fully contain the target. Position anomalies or slight drifts are corrected, if necessary. As illustrated in Fig. 1, the content of theses bounding boxes is extracted, rotation-normalized (each vehicle is moving in the right direction) and resized to 32 × 16 pixels. The extracted vehicles are further stored in single image files together with their meta information such as frame ID, vehicle ID and position in the cropped frame.

The focus of this work lies on evaluating the performance of appearance based descriptors in order to solve the problem of vehicle re-identification using the mentioned dataset. We generate closed sets of query vehicles and potential candidates. A query is here defined as a vehicle located in both frames of a pair. In this way, we prevent scenarios in which a vehicle located in the first frame of the pair has to be re-identified although it is not present in the second frame. These closed sets are generated by searching ξ candidates in the second frame next to the query vehicle’s position in the first frame. This approach is inspired by the
4. Appearance Based Vehicle Re-identification in WAMI Data

Several different descriptors with specific properties useful for person re-identification have already been proposed in the literature [9, 35]. Nevertheless, the field of vehicle re-identification in WAMI data based on appearance features is a special case that has not been discussed sufficiently in the literature so far. In general, there are three categories of descriptors: texture based, shape based and color based descriptors [42]. However, since our dataset is based on monochromatic images, color based features will not be considered. Furthermore, the descriptors should be invariant to major changes in illumination due to rapidly varying camera gain and minor changes in orientation and appearance due to the low frame rate and slight shifts of the objects inside the bounding boxes as a result of imprecise detection. Some example vehicles are illustrated in Fig. 1.

The gray value histogram is a first simple descriptor that is able to extract the vehicle appearance information. However, there are more powerful descriptors that consider texture, shape, or covariance. We identified several suitable descriptors and introduce them in the remainder of this section. Each descriptor is evaluated in Section 5.

LBP is a highly discriminative texture operator. It is based on summarizing the local structure in an image by thresholding the 3 × 3-neighborhood of each pixel with the center pixel value and labeling the result as a binary number [23]. A texture descriptor can be created by generating the histogram of theses labels. The key advantage of this descriptor is its invariance to monotonic gray-level changes and to rotation [24]. Local contrast can be used as additional information by calculating the Local Variance Measure (VAR) [24] histogram. LBP and VAR can then be combined by simple concatenation. This fused histogram is denoted by LBP / VAR [24]. In our evaluation, the rotation-invariant uniform $LBP_{P,R}^{ui}$ / VAR$_{P,R}$ descriptor is considered, where $P$ is the number of neighboring pixels used for LBP calculation and $R$ is the radius of the neighbors w.r.t. the center pixel. The descriptor is denoted by $H_{lbpvar}$, where $H$ represents a histogram. By dividing the image into smaller regions (grid cells) and building the region-specific histograms separately, the spatial information can be preserved by concatenating the region-specific histograms [2]. In our experiments, we subdivide the image into 8 non-overlapping regions of size $8 \times 8$ pixels, and use the notation $H_{local(lbpvar)}$. In order to avoid histogram sparsity, $LBP_{P,R}^{ui}$ and VAR$_{P,R}$ are calculated multiple times with a fixed number of $P = 8$ and varying radius $R$. We use $R \in \{1, 2\}$.

Since the vehicle texture may not contain enough information for vehicle re-identification, the shape can be taken into consideration. In order to describe the shape information we use the Histogram of oriented Gradients (HoG) that was originally designed as shape descriptor for human detection [14]. This descriptor uses the fact that local object appearance and shape within an image can be described by the distribution of the intensity gradient. In our experiments, the HoG is calculated for $8 \times 8$ pixels overlapping blocks, with cells of $4 \times 4$ pixels and a block-stride of 4 pixels.

The descriptors introduced so far are histograms. However, when combining several different features in a joint representation by histogram concatenation, the dimension of the combined descriptor grows exponentially with the number of features. Tuzel et al. [38] proposed to reduce the dimensionality by using the covariance of several image statistics computed inside a region of interest as a region descriptor. For using covariance matrices as descriptors, features such as gray value, LBP, VAR, or gradient magnitude are first determined for each pixel of the image. Then, $c$-dimensional feature vectors are constructed for each pixel. For a given image $I$, let $\{z_k\}_{k=1 \ldots n}$ be the $c$-dimensional feature points in $I$. The image is represented by the $c \times c$ covariance matrix of the feature points

$$C_I = \frac{1}{n-1} \sum_{k=1}^{n} (z_k - \mu)(z_k - \mu)^T,$$  \hspace{1cm} (1)

where $\mu$ is the mean of the points [38].

We first consider $Cov(I_g)$ by using a three-dimensional feature vector generated from the pixel locations $(x, y)$ and
the gray values \( I(x, y) \):

\[
F_{I_g}(x, y) = \begin{bmatrix} x & y & I(x, y) \end{bmatrix}^T
\]  

Then, a gray value image \( I_g \), an LBP image \( I_{lbp} \) and a VAR image \( I_{var} \) are combined in order to build a five-dimensional feature vector:

\[
F_{I_g I_{lbp} I_{VAR}}(x, y) = \begin{bmatrix} x & y & I(x, y) & I_{lbp}(x, y) & I_{VAR}(x, y) \end{bmatrix}^T
\]  

The additional LBP and VAR images are both computed for \( P = 8 \) and \( R = 1 \).

\[
F_{I_g I_{lbp} I_{VAR}}(x, y) = \begin{bmatrix} x & y & I(x, y) & I_{lbp}(x, y) & I_{VAR}(x, y) \end{bmatrix}^T
\]  

Finally, a seven-dimensional feature vector integrating all mentioned information is used for computing \( Cov(I_g I_{lbp} I_{VAR} I_m I_\theta) \).

The matching is performed by minimizing the matching distance \( d \) between the query and the candidates. In the context of this paper, the objective is to obtain a ranking based on these distances. Additionally to the Euclidean distance, the Hellinger distance is considered, since it has been demonstrated in areas such as texture classification and image categorization, that using Euclidean distance often yields inferior performance compared to Hellinger distance [5]. We also evaluate the \( \chi^2 \) distance because of its promising result for LBP in [1]. Considering that the covariance matrices do not lie on Euclidean space [7], the just mentioned distances for histogram based descriptors are not suitable for covariance based descriptors [38]. Therefore, the covariance manifold is often specified as Riemannian in order to use powerful tools from differential geometry [28] such as the geodesic distance \( p(x_i, x_j) \) [6]. It is relevant to note that an equivalent form of the geodesic distance can be given in terms of generalized eigenvalues [16], which is used in this work.

5. Experimental Setup and Results

5.1. Evaluation Measures

The most important evaluation measure is the rank-1 Recognition Rate (RR) that represents the percentage of correct matches at rank one of the ranking. However, since RR does not provide sufficient information about the overall effectiveness of a descriptor for re-identification tasks, we use the Cumulative Matching Characteristic (CMC) to evaluate our results. This metric shows the cumulative probability of correctly re-identifying a target at rank \( r \) or below for a closed set and is commonly used for person re-identification [9]. A point to note about the CMC curve is that the curve slope depends on the number of candidates. Thus, the CMC curve is always lower with a larger number of candidates [20]. Therefore, for most of our experiments the number of candidates for a query is fixed to \( \xi = 10 \) candidates. We also calculate the normalized Area under the CMC Curve (nAuC) which provides the probability of a correct match [9].

5.2. Results

The candidates’ ranking is first calculated for each query in each pair of frames. The overall matching statistic is then averaged over all pairs of frames by calculating mean \( \mu_{rank} \) and standard deviation \( \sigma_{rank} \). Histogram based approaches are abbreviated by \( H \). For example, \( H_g \) represents the gray value histogram or \( H_{lbp} \) stands for the LBP histogram. Furthermore, \( H_{local(lbpvar)} \) represents the Local LBP \( P,R \)/VAR \( P,R \) histogram. Analogously to the histogram based approaches, the covariance based approaches are abbreviated as \( Cov(\cdot) \) where \( \cdot \) are the feature images introduced in Section 4.

We first compare the Euclidean distance, the Hellinger distance and the \( \chi^2 \) distance for three histogram based descriptors: the gray value histogram \( H_g \) is used as a baseline and compared to the texture based region descriptor \( H_{local(lbpvar)} \) and the shape based region descriptor HoG. The results in Fig. 3 confirm the conclusions of Arandjelovic and Zisserman [5] that the Hellinger distance outperforms the Euclidean and the \( \chi^2 \) distance. Hence, for all other experiments we use the Hellinger distance.

The results for all descriptors with \( \xi = 10 \) candidates for each query are presented in Table 1. We calculated CMC curves for each descriptor and display the nAuC in Table 1 for a compact presentation of our results. The best descriptor performance is displayed in red color.

For a better visualization of the result, only chosen texture based descriptor with \( \xi = 10 \) candidates for each query
are shown in Fig. 4. All descriptors show stable performance except for \( H_g \) with a standard deviation of 18 percent at rank 1. The performance of the covariance descriptor increases with more information included. However, Table 1 shows only minor improvements from the \( \text{Cov}(I_g I_{lbp} I_{var} I_m I_\theta) \) to \( \text{Cov}(I_g I_{lbp} I_{var} I_m I_\theta) \), indicating an upper limit. Although \( \text{Cov}(I_g I_{lbp} I_{var} I_m I_\theta) \) considers more information compared to the shape based HoG descriptor, they perform at the same level. The \( H_{local(lbpvar)} \) region histogram, which produces over 30 percent better results than its global variant, shows the most promising results with about 5 percent standard deviation and a recognition rate over 80 percent at rank one.

Figure 6 shows the CMC curves for varying values of \( \xi \). For better readability the variation of \( \xi \) for each descriptor is superscripted, e.g. \( H_{local(lbpvar)}^{\xi} \), \( \text{HoG}^{\xi} \), \( \text{Cov}(I_g I_{lbp} I_{var} I_m I_\theta)^{\xi} \) and the standard deviation is omitted. As expected, the performance of all descriptors decreases when the number of targets increases. However, we can see that with increasing \( \xi \), the performance of \( H_{local(lbpvar)}^{\xi} \) decreases significantly slower than the performance of the two other descriptors. Furthermore, for all values of \( \xi \) an accumulated recognition rate of 100 percent is reached at a lower number of considered candidates.

6. Conclusions

In this paper, we systematically evaluated appearance based vehicle descriptors for WAMI data processing. We modified the GT of the WPAFB 2009 dataset in order to generate a set of 6,968 queries of a vehicle re-identification ranking problem. Due to the low resolution of individual vehicles and the lack of color information in the WPAFB 2009 dataset, we identified and evaluated several suitable descriptors based on texture, shape, and covariance. We propose to use the combination of LBP and VAR histograms that are calculated in local grid cells of the image as descriptor and the Hellinger distance as measure for matching similarity. Our results can be used to improve appearance based multiple object tracking algorithms and real-time database search algorithms.
Table 1: Recognition results for all descriptors evaluated on our WPAFB 2009 subset with 6,968 queries using the Hellinger distance and $\xi = 10$ candidates for each query.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>RR</th>
<th>$\sigma_{RR}$</th>
<th>nAuC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_g$</td>
<td>0.588</td>
<td>0.189</td>
<td>0.836</td>
</tr>
<tr>
<td>$H_{lbp}$</td>
<td>0.403</td>
<td>0.066</td>
<td>0.771</td>
</tr>
<tr>
<td>$H_{var}$</td>
<td>0.398</td>
<td>0.055</td>
<td>0.772</td>
</tr>
<tr>
<td>$H_{lbpvar}$</td>
<td>0.506</td>
<td>0.066</td>
<td>0.802</td>
</tr>
<tr>
<td>$H_{local(lbp)}$</td>
<td>0.827</td>
<td>0.049</td>
<td>0.918</td>
</tr>
<tr>
<td>$H_{local(lbpvar)}$</td>
<td>0.848</td>
<td>0.049</td>
<td>0.923</td>
</tr>
<tr>
<td>$HoG$</td>
<td>0.782</td>
<td>0.065</td>
<td>0.895</td>
</tr>
<tr>
<td>$Cov(I_g)$</td>
<td>0.653</td>
<td>0.067</td>
<td>0.872</td>
</tr>
<tr>
<td>$Cov(I_gI_{lbp})$</td>
<td>0.721</td>
<td>0.062</td>
<td>0.888</td>
</tr>
<tr>
<td>$Cov(I_gI_{lbpI_{var}})$</td>
<td>0.740</td>
<td>0.054</td>
<td>0.893</td>
</tr>
<tr>
<td>$Cov(I_gI_{lbpI_{varI_{mI_{a}}}})$</td>
<td>0.760</td>
<td>0.053</td>
<td>0.898</td>
</tr>
<tr>
<td>$Cov(I_gI_{mI_{a}})$</td>
<td>0.754</td>
<td>0.050</td>
<td>0.896</td>
</tr>
</tbody>
</table>

References


