In this paper, we propose a new image prior customized in average forward motion images - average image prior to decompose reflection and obstruction layers from in-vehicle black box videos. The average image prior is a statistic observation of the average in-vehicle video frames. This prior is derived from one simple observation that most gradients in the background of average forward motion images show a same direction toward a vanishing point in the average image. Base on this prior, not only for validation of the reflection layer decomposition but also we demonstrate that the proposed method can be effectively suitable for the obstruction such like dirt, dust on a windscreen and dashboard-free images.

1. Introduction

The popularity of in-vehicle black box cameras (dashboard camera) has been explosively increased up with importance to the security in many countries. However, studies only focus on degradation factors such as the windscreen reflection, dirt, dust and dashboard in black box contents still have not been fully considered following with the increment. In the conventional layer separation model, the observed image $I$ is composed with reflection layer $R$ and background layer $L$.

$$I = R + L$$  \hspace{1cm} (1)

This model is fully ill-posed, even for a number of different mixture images. Previous methods (Yu and Brown’s [7], Levin [1]) studied on conventional images can be applied the contents to decompose a reflection layer but their results can’t reach to sufficient visualization and suffer from inconsistency of tone respectively. In addition, to address the readability issue, our problem model should have one more obstruction layer $O$. Thus final problem model is that the observed image $I$ is composed with the reflection layer $R$, the background layer $L$ and an obstruction layer $O$.

$$I = R + L + O$$  \hspace{1cm} (2)

To address the problem, we investigate the reflection and obstruction in the record videos to find distinct properties from the conventional imaging condition applied to previous works. That is, the reflection and the obstruction scenes tend to be stationary while the background scene keeps rapidly on moving along the direction of vanishing lines. This particular circumstance in the in-vehicle black box video has inspired the effective approach proposed in this paper. In following chapters, we demonstrate the efficiencies of our method via statistic validation of the average prior and the final results.

2. Proposed Algorithm

The reflection and the obstruction on a glass surface can be deduced as a layer separation model. The proposed model in this work is inspired by the layer separation model in [1], in which the joint probability of two layers $L$ and $R$ should be maximized: $Pr(L,R) = Pr(L)Pr(R)$. From this model and the observation that the average of consecutive frames $\tilde{I}$ in a video where $\tilde{I} = \overline{L + R + O}$ because the reflection and obstruction on the windscreen are stationary $R' + O' \approx R'_{\text{avg}} + O'_{\text{avg}} = \overline{R + O}$. It is possible to omit the time domain subscript and regard the reflection and obstruction layers as a single constant stationary layer $S = R + O$. Therefore, the complexity of three layer separation model can be simplified to two layer separation model by integration of the reflection and obstruction prior distributions as follows.

$$E(L, R, O) = E(L, S) = \sum_{\{i,N\}} \sum_{j} F_1(D^j \tilde{I}_i - D^j S_i) + \lambda F_2(D^j S_i) \hspace{1cm} (3)$$

$\tilde{I}$ is a constant over several frames taken to obtain the average image and $\lambda$ denotes weight parameter. Stepping from (3), proper prior distribution model for $F_1$ and $F_2$ should be imposed to perform reliable layer separation. The detail of the average image prior and the reflection and obstruction removal method are described in the following subsections.

2.1 Average Image Prior

The average image prior is based on the observation on the in-vehicle videos: the background layer keeps on moving along the direction of vanishing lines. In other words, the most gradients in the background of the average image show a same direction toward a vanishing point. To verify the prior, we collect black box videos on the
internet and construct 155 average images by 100~150 frames. Since most videos are recorded in-vehicle, we manually cropped patches on the reflection and obstruction-free images. To represent the average background image for each patch, we set the patch size 100x100 which is larger than conventional patch-based methods. Fig. 1. shows several average images and gradient distributions filtering on the polar coordinate for RGB color. The polar coordinate gradient distributions shows short-tail distribution which models narrow-Gaussian (\( \text{Pr}(x) = e^{-x^2/\sigma^2} \)).

Fig. 1. Average image dataset. (a) Example average images from black box video collections. (b) Gradient distribution on Polar coordinates in each color channel (short-tail).

2.2 Relative Sparsity Distribution using Region Division

In the proposed approach, the average image is divided into several different angular regions to utilize different shape of hyper-Laplacian distributions with different \( \alpha \) values. We divide the average image into several angular regions by the vanishing point to utilize different shapes of hyper-Laplacian with different \( \alpha \) values. To determine \( \alpha \) value for each region, detect edge pixels in the stationary layer by via edge frame \( E_i \) applied Sobel edge detector (threshold: 0.01).

\[
E_s(x) = \begin{cases} 
1, & \text{Sig} \left( \frac{\sum_{n \in E_i(x)} E_s(n)}{N} \right) > 0.95 \\
0, & \text{else} 
\end{cases}
\]

and then determine the \( \alpha \) value by thresholds from three candidate of hyper-Laplacian distribution models: \( \beta_{90,0,70,0} \) (0.01 \( \leq x \) = [ 0.1103, 0.0577, 0.0285 ]).

\[
P_{\alpha_{90,0,70,0}} (0.01 \leq x) = \frac{\sum_{n \in E_s(n)} E_s(n)}{N_{\alpha_{90,0,70,0}}} \] (4)

\[
\alpha_{90,0,70,0} = \begin{cases} 
0.6, & P_{\alpha_{90,0,70,0}} > 0.11 \\
0.7, & 0.11 \geq P_{\alpha_{90,0,70,0}} > 0.058 \\
0.8, & \text{otherwise} 
\end{cases}
\]

where \( \text{Sig}(x) \), \( \alpha_{\text{Re}_i} \) and \( N_{\text{Re}_i} \) denote sigmoid function, an alpha value and a number of pixels in for Region \( i \) respectively.

2.3 Reflection and Obstruction Removal

The final objective function with non-negativity constraints is as follows

\[
\min_{S} \sum_{i \in N} \sum_{j \in J_i} (D_j^S - D_j^S)^2 + \lambda_{E_i} \sum_{h \in \text{Bin}}(\sum_{n \in E_i} \sum_{j \in J} \left| D_j^S \right|^{2\alpha}) + \lambda_p \min(0, (\bar{T} - S)^2) + \min(0, S^2))
\]

To accelerate the optimization, we use coarse-to-fine method across multi-scale images (small-to-large) built by image pyramid for the input frame.

2.4 Post-Processing

In practical circumstances, the stationary property over all frames would not be realistic due to slight vibrations by environment factors. To mitigate the circumstances, we adopt fine motion-flow \( u \) and find optimal it defined as

\[
\min_{u} \sum_{n \in N} \left| V(L_n - W(u)S) \right| + \lambda (\nabla u)^2
\]

where \( W(u) \) is a warping matrix.

After optimization, the background scene can be decomposed by

\[
L_t = I_t - W(u)S
\]

For \( L_t \), we use the angular region based luminance transfer method [4] with the input frame as a reference image. The efficiency of fine motion flow is proved by shown as Fig. 2.

Fig. 2. Comparison proposed method with and without the fine motion flow.

3. Experimental Results

To demonstrate the robustness of the proposed method, we compare the performance of reflection removal with the state-of-the-art algorithms by Yu and Brown [7] and Levin [1] while BM3D [2] and Eigen [3] are compared to show the performance of obstruction removal. In Fig. 3 and Fig 4, the proposed method shows better performance in removing the reflection and the obstruction of the real-
world videos.
In order to validate the robustness of the proposed algorithm in tough environment quantitatively, we test a few synthetic frames including reflection and obstruction factors using CamVid dataset [5, 6]. In Table 1, PSNR is calculated for each selected frame to prove that the proposed method has high rate than those of the previous methods in [2, 3, 7].

![Fig. 3. Frame by frame results with comparison (reflection).](image)
The first row shows the average image from the input frames and the result of separated reflection layers from the initial sample frame. (a) Input frames. (b) Results of [1]. (c) Results of [7]. (d) Results of the proposed method.

![Fig. 4. Frame by frame results with comparison (dirt and reflection).](image)
(a) Input frames. (b) Result of [2]. (c) Results of [3]. (d) Results of [7]. (d) Results of the proposed method.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>PSNR</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>seq06R0</td>
<td>19.26</td>
<td>17.19</td>
</tr>
<tr>
<td>seq16E5</td>
<td>17.66</td>
<td>15.87</td>
</tr>
<tr>
<td>seq05VD</td>
<td>13.58</td>
<td>11.42</td>
</tr>
<tr>
<td>BM3D [2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dirt removal [3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yu and Brown [7]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed method</td>
<td>19.57</td>
<td>17.80</td>
</tr>
</tbody>
</table>

4. Conclusion
In this paper, we propose the reflection and obstruction removal method to enhance readabilities of in-vehicle black box cameras.

5. Acknowledgement
This work was supported by the IT R&D program of MSIP/KEIT. [10047078, 3D reconstruction technology development for scene of car accident using multi view black box image].

6. Reference