Escaping Specularity: Recovering Specular-Free Video Sequences from Rank-Constrained Data

Samar M. Alsaleh
The George Washington University
sm57@gwu.edu

Angelica I. Aviles
Universitat Politècnica de Cataluya
angelica.ivone.aviles@upc.edu

Alicia Casals
Universitat Politècnica de Cataluya
alicia.casals@upc.edu

James Hahn
The George Washington University
hahn@gwu.edu

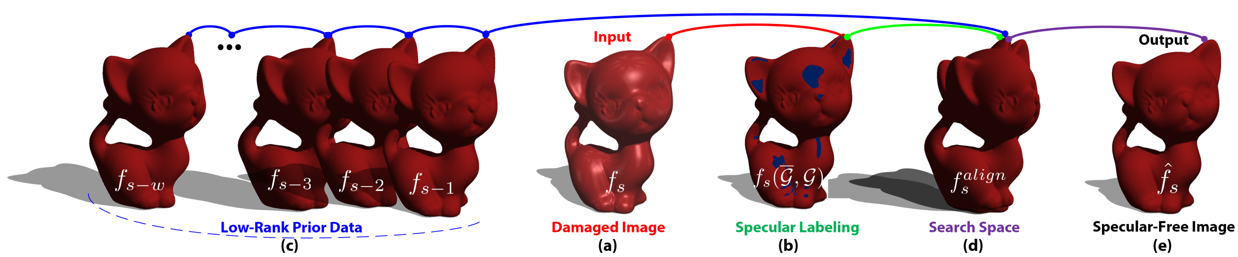

Figure 1: Recovering specular-free video sequences from rank-constrained data. (a) an input image from the sequence. (b) labeling results of the damaged regions. (c) low-rank representation of prior temporal data. (d) search space constructed from aligning the prior data and used to find the optimal information for the missing region. (e) resulted specular-free output.

CCS CONCEPTS

• **Computing methodologies** → **Image processing**; *Computer graphics*; *Image manipulation*;

KEYWORDS

Specular highlights, Low-rank data, Sparse alignment

ACM Reference format:
Samar M. Alsaleh, Angelica I. Aviles, Alicia Casals, and James Hahn. 2017. Escaping Specularity: Recovering Specular-Free Video Sequences from Rank-Constrained Data. In *Proceedings of SIGGRAPH '17 Posters, Los Angeles, CA, USA, July 30 - August 03, 2017,* 2 pages.
https://doi.org/10.1145/3102163.3102235

1 INTRODUCTION

The appearance of objects is significantly affected by the illumination conditions in the environment. Particularly with objects that have strong reflectivity as they suffer from more dominant specular highlights, causing information loss and discontinuity in the image domain. Many computer vision algorithms are vulnerable to errors in the presence of specular highlights because they violate the image consistency assumption and hinder the performance of many vision tasks, such as object recognition, tracking and surface reconstruction [Artusi et al. 2011]. This is further complicated when we consider video sequences with free-moving cameras or dynamic objects, which is the focus of this work.

Most existing solutions for retrieving missing information in complex video sequences, for example [Ebdelli et al. 2015; Newson et al. 2014], are based on a set of minimization procedures, sometimes up to four, which results in a high computational demand. They also use temporal information and align neighboring frames using homography-based algorithms, which is prone to failure, causing error propagation and unpleasant visual results.

We present a novel framework for specular-free video recovery that is equally effective for both static and moving camera and with the presence of object motion (see Figure 1). Our main contributions and what distinguish our work from existing approaches are:

- A fast and effective labeling process based on a set of color variation and gradient information conditions.
- A low-rank representation of prior temporal data which decreases computational time and results in better minima.
- A robust sparse-based alignment solution that reduces error propagation.

2 TECHNICAL APPROACH

Consider a video sequence $F = \{f_s\}_{s=1}^{S}$ with S frames of size $\mathcal{X} \times \mathcal{Y}$ and a finite Lipchitz domain $\Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^d$ with $d = 3$ for the red, green, and blue (RGB) channels. To recover specular-free frames, our solution applies two main steps illustrated in Figure 1. The first step is an adaptive specular highlights labeling approach that takes into account information from the current frame to accurately label

SIGGRAPH '17 Posters, July 30 - August 03, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5015-0/17/07.
https://doi.org/10.1145/3102163.3102235

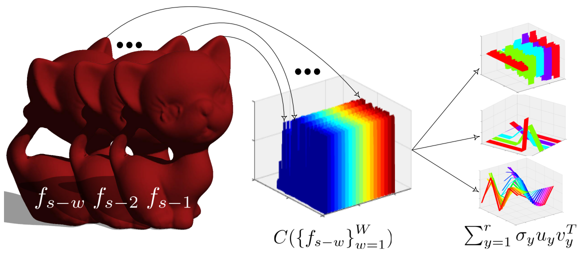

Figure 2: Our solution arranges the prior temporal data in a Casorati Matrix C in order to explode the correlation between the data and find the low r rank.

specular pixels. Two labels are used in this step: (i) specular region ($\overline{\mathcal{G}}$) and (ii) non-specular region ($\mathcal{G}$).

A given pixel $f_s(x, y)$ is labeled as $\overline{\mathcal{G}}$ if it meets one of two conditions. The first condition identifies high intense specular regions using a color-adaptive technique that takes into account the standard deviation of the color variations in the given frame. Pixels that fall within one standard deviation away from the brightest pixel value (the one with maximum intensity) are considered specular pixels. The second condition uses the gradient information to detect the less intense parts of the specular highlights.

Once we have the labeling results, the second part of the solution is a robust inpainting process that utilizes low-rank prior temporal data to restore the damaged regions. Inpainting finds the optimal value for each missing pixel in $\overline{\mathcal{G}}$ by minimizing an energy functional that uses two data sources: the labeling results and a low-rank search space.

The search space f_s^{align} is created by aligning a short-term temporal window of W past frames represented in a low-rank manner and denoted as $\overline{F}_W = \{f_{s-w}\}_{w=1}^{W}$. The low-rank representation of the frames is achieved using a Casorati matrix and applying the theorem of decomposition to exploit the high correlation between the consecutive frames (see Figure 2). The alignment of the low-rank frames in $\overline{F}_W$ is achieved using an energy functional that finds the best transformation that aligns a set of control points, initially uniformly spaced, defined on the frames.

3 EXPERIMENTAL RESULTS

We demonstrate the performance of our approach by carrying out experimentation with different synthetic and in-vivo datasets. All results and comparisons were run under the same condition using an Intel(R) Core i7-32GB RAM, and a Nvidia GeForce GT 610.

We quantitatively evaluated our labeling results against the ground truth using three measures: Dice's coefficient Dc, which measures the similarity between two labeling results, accuracy, and error rate. For all the datasets, the overall averages were 0.82, 99% and 1.94% for Dc, accuracy and error rate respectively.

To evaluate the inpainting, we compared our approach against two commercially available softwares (Inpaint software [1] and adobe Photoshop [2]) and also against Newson's solution [Newson et al. 2014] which is, to our best knowledge, one of the most robust approaches. We used the homogeneity variation in our evaluation

[1]https://www.theinpaint.com/
[2]https://www.adobe.com/Photoshop/

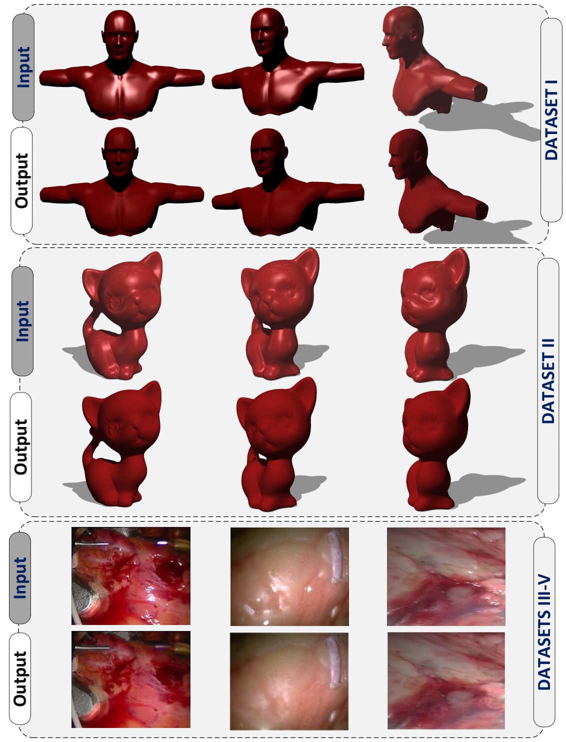

Figure 3: Evaluation was performed exhaustively in five datasets and the results show visually pleasing, specular-free outputs even with complex scenes.

measured by the coefficient of variation CV. The results showed that our approach outperformed the compared approaches and achieved the smallest CV for all datasets.

We also evaluated the repercussion of promoting low-rank in our solution in terms of computational time. The results showed that promoting low rank decreased the computational time to an average of 6 times less than using full-rank data. In average, full-rank demanded 23 iterations per frame while low-rank 18 iterations. This shows that promoting low-rank allowed our global approach to achieve a good minima in a computationally efficient manner.

Apart from the numerical results, our global approach can be visually inspected in Figure 3 where we selected interesting frames that show complex cases. Our approach was able to recover the lost information and achieve visually pleasing results with those cases and even with complex objects, such as organs (heart dataset), where our solution was able to retrieve the texture accurately.

REFERENCES

Alessandro Artusi, Francesco Banterle, and Dmitry Chetverikov. 2011. A survey of specularity removal methods. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 2208–2230.

Mounira Ebdelli, Olivier Le Meur, and Christine Guillemot. 2015. Video inpainting with short-term windows: application to object removal and error concealment. *IEEE Transactions on Image Processing* 24, 10 (2015), 3034–3047.

Alasdair Newson, Andrés Almansa, Matthieu Fradet, Yann Gousseau, and Patrick Pérez. 2014. Video inpainting of complex scenes. *SIAM Journal on Imaging Sciences* (2014).