Multi-View Separation of Background and Reflection by Coupled Low-Rank Decomposition

Jian Lai, Wee Kheng Leow, Terence Sim, and Guodong Li

Dept. of Computer Science, National University of Singapore
laij, leowwk, tsim@comp.nus.edu.sg, lguodong@u.nus.edu

Abstract. Images captured by a camera through glass often have reflection superimposed on the transmitted background. Among existing methods for reflection separation, multi-view methods are the most convenient to apply because they require the user to just take multiple images of a scene at varying viewing angles. Some of these methods are restricted to the simple case where the background scene and reflection scene are planar. The methods that handle non-planar scenes employ image feature flow to capture correspondence for image alignment, but they can overfit resulting in degraded performance. This paper proposes a multiple-view method for separating background and reflection based on robust principal component analysis. It models the background and reflection as rank-1 matrices, which are decomposed according to different transformations for aligning the background and reflection images. It can handle non-planar scenes and global reflection. Comprehensive test results show that our method is more accurate and robust than recent related methods.

Keywords: Reflection removal, non-planar scenes, robust PCA,

1 Introduction

When an image is captured by a camera through a piece of glass, reflection is often superimposed on the transmitted background scene. Reflection is annoying because it corrupts the image content, and it is difficult to remove them manually. Therefore, separation of background and reflection from the superimposed image has attracted research interests over the years. Existing methods for reflection separation can be broadly grouped into three categories: single-image, single-view multiple-image, and multi-view multiple-image. Single-image methods [12, 13] in general require user inputs or training images, which are inconvenient to apply. Single-view methods [1, 6, 10, 21, 22] take multiple images of a scene with a fixed camera under varying imaging conditions such as polarization, flash/no-flash, and focusing, and use the resulting differences between background and reflection to separate them. These methods require additional accessories, professional setting skill, and fixed camera position, and so are not suitable for images taken by non-professional users.

Multi-view methods take multiple images of a scene at varying viewing angles. Thus, the images are not aligned. The reflection of objects as seen by a camera, which we call *virtual scene*, is located on the same side of the reflecting glass as the background scene. When the background scene and the virtual scene are not coplanar, the

background image and reflection image can be separated by multi-view methods. When the objects in the background are on a plane, and similarly for the reflection, the background and reflection scenes are planar and can be aligned by single homography. Several methods [7, 8, 23] assume this simplifying condition. Some recent methods [14, 24, 25] try to solve the more complex case of non-planar scenes. They employ nonlinear flows of image features to capture correspondence for image alignment, but they can overfit resulting in degraded performance [16].

This paper proposes a multiple-view method for separating reflection and background by coupled low-rank decomposition (CLORD). It is based on the method of robust principal component analysis (robust PCA) [4], which has been successfully applied to various computer vision problems [20, 11, 18, 19]. It models the background and reflection as rank-1 matrices, which are decomposed according to different transformations for aligning the background and reflection images. It can handle non-planar scenes and global reflection. Comprehensive test results show that CLORD is more accurate and robust than recent related methods.

2 Coupled Low-Rank Decomposition

2.1 Overview

Given a set of superimposed images of stationary objects, which are captured over a small range of viewing angles, CLORD is proposed to recover the background and foreground simultaneously. CLORD is inspired by but differs from that of Guo et al. [8]. In comparison, Guo et al. [8], Szeliski et al. [23], and Gai et al. [7] assume planar scenes, and Guo et al. [8] also assume sparse reflection. So, their methods solve a simpler version of the reflection separation problem. Like CLORD, Li and Brown [14], Xue et al. [24] and Yang et al. [25] also handle non-planar scenes and global reflection. However, CLORD is based on robust PCA [4], which has been successfully applied to various challenging computer vision tasks, whereas [14], [24] and [25] apply iteratively reweighted least squares.

2.2 Problem Formulation

For a set of superimposed images $\mathbf{f}'_i$ with m pixels, $i = 1, \ldots, n$, captured at different viewing angles, each image $\mathbf{f}'_i$ is a linear combination of the transmitted background $\mathbf{b}'_i$ and the reflection $\mathbf{r}'_i$, i.e., $\mathbf{f}'_i = \mathbf{b}'_i + \mathbf{r}'_i$. By arranging each image as a column in an $m \times n$ matrix, the above relationship can be written as $\mathbf{F}' = \mathbf{B}' + \mathbf{R}'$, where $\mathbf{F}'$, $\mathbf{B}'$, and $\mathbf{R}'$ denote the matrices of the unaligned superimposed images, transmitted background, and reflection, respectively. The corresponding matrices $\mathbf{B}$ and $\mathbf{R}$ of aligned background and reflection are low-rank, specifically, rank-1. They are related to the unaligned matrices by transformation functions T and T':

$$\mathbf{B} = T_b(\mathbf{B}'), \;\; \mathbf{R} = T_r(\mathbf{R}'), \;\; \mathbf{B}' = T'_b(\mathbf{B}), \;\; \mathbf{R}' = T'_r(\mathbf{R}).$$

So, $\mathbf{F}' = T'_b(\mathbf{B}) + T'_r(\mathbf{R})$. The transformations T and T' differ for different $\mathbf{b}'_i$ and $\mathbf{r}'_i$. They are nonlinear in general, and T' is a good approximation of T^{-1}.

Distinctive features such as edges correspond to large gradients in the images. Image gradients are computed by convolving gradient filter kernel $\mathbf{g}$ with the background

and the reflection giving $\mathbf{g} * \mathbf{b}_i$ and $\mathbf{g} * \mathbf{r}_i$. Since the gradient filter is linear, these convolutions can be written in matrix form as $\mathbf{DB}$ and $\mathbf{DR}$, where $\mathbf{D}$ is gradient operator.

To enforce mutual exclusion of edges, two coupled weight matrices $\mathbf{W}_b$ and $\mathbf{W}_r$ are applied to the gradients giving $\mathbf{W}_b \circ \mathbf{DB}$ and $\mathbf{W}_r \circ \mathbf{DR}$, where $\circ$ denotes element-wise multiplication. If an edge belongs to the background, its corresponding entry in $\mathbf{W}_b$ is small whereas its entry in $\mathbf{W}_r$ is large. The converse is true if edge belongs to reflection.

Now, the reflection separation problem can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{B},\mathbf{R},\mathbf{E}',\mathbf{N}',T_b',T_r'} \; & \|\mathbf{B}\|_* + \|\mathbf{R}\|_* + \lambda_1\|\mathbf{E}'\|_1 + \frac{\lambda_2}{2}\|\mathbf{N}'\|_F^2 + \\ & \lambda_3\|\mathbf{W}_b \circ \mathbf{DB}\|_1 + \lambda_3\|\mathbf{W}_r \circ \mathbf{DR}\|_1 \\ \text{subject to } \; & \mathbf{F}' = T_b'(\mathbf{B}) + T_r'(\mathbf{R}) + \mathbf{E}' + \mathbf{N}'. \end{aligned} \tag{1}$$

The first two terms are the nuclear norms of background and reflection, which are small for low-rank matrices. The third term uses l_1-norm to model sparse noise $\mathbf{E}'$ with possibly large amplitudes whereas the fourth term uses Frobenius norm to model non-sparse, small-amplitude noise $\mathbf{N}'$. The last two terms model mutually exclusive sparse gradients. This formulation handles the aligned background $\mathbf{B}$ and aligned reflection $\mathbf{R}$ in a uniform manner. This uniformity results in a simpler objective function compared to that of Guo et al. [8], which models low-rank background with sparse reflection instead. Moreover, they model mutual exclusion of features with $\|\mathbf{DB} \circ \mathbf{DR}\|_1$. The minimization of $\|\mathbf{DB} \circ \mathbf{DR}\|_1$ can be achieved by small $\mathbf{DB}$ and small $\mathbf{DR}$, which is ambiguous where the edge belongs. In comparison, CLORD is more specific and less prone to such ambiguity (Section 2.5).

In general, the background scene and the virtual scene are not coplanar. So, they have different transformations T_b and T_r, which permit $\mathbf{B}$ and $\mathbf{R}$ to be separated. In the degenerate case that the background and virtual scenes are coplanar, separation of background and reflection is impossible regardless of the method used.

To solve Problem 1, an alternating optimization algorithm is applied to minimize various parts of the objective function alternatively. The details of our algorithm CLORD is derived in a similar manner as [8, 9, 11, 15, 19, 20].

2.3 Optimization of Background

First, let us consider the optimization of $\mathbf{B}$, $\mathbf{E}'$, and $\mathbf{N}'$ while keeping $\mathbf{R}$, T_b', and T_r' fixed. As the superimposed images are captured at a small range of varying viewing angles, the structures of the images are essentially the same. So, minimizing $\mathbf{E}'$ and $\mathbf{N}'$ is equivalent to minimizing $\mathbf{E} = T_b(\mathbf{E}')$ and $\mathbf{N} = T_b(\mathbf{N}')$. Denoting $\mathbf{F} = T_b(\mathbf{F}') - T_b(T_r'(\mathbf{R}))$ and applying T_b to both sides of the constraint equation of Problem 1 yields equivalent constraint $\mathbf{F} = \mathbf{B} + \mathbf{E} + \mathbf{N}$. With $\mathbf{R}$, T_b', and T_r' fixed, Problem 1 reduces to

$$\begin{aligned} \min_{\mathbf{B},\mathbf{E},\mathbf{N}} \; & \|\mathbf{B}\|_* + \lambda_1\|\mathbf{E}\|_1 + \frac{\lambda_2}{2}\|\mathbf{N}\|_F^2 + \lambda_3\|\mathbf{W}_b \circ \mathbf{DB}\|_1 \\ \text{subject to } \; & \mathbf{F} = \mathbf{B} + \mathbf{E} + \mathbf{N}. \end{aligned} \tag{2}$$

To match Problem 2 to robust PCA's formulation, we introduce two auxiliary variables $\mathbf{A}$ and $\mathbf{C}$ such that $\mathbf{B} = \mathbf{A}$ and $\mathbf{DB} = \mathbf{C}$. Then, Problem 2 becomes

$$\begin{aligned} \min_{\mathbf{B},\mathbf{E},\mathbf{N},\mathbf{A},\mathbf{C}} \; & \|\mathbf{A}\|_* + \lambda_1\|\mathbf{E}\|_1 + \frac{\lambda_2}{2}\|\mathbf{N}\|_F^2 + \lambda_3\|\mathbf{W}_b \circ \mathbf{C}\|_1 \\ \text{subject to } \; & \mathbf{F} = \mathbf{B} + \mathbf{E} + \mathbf{N},\ \mathbf{B} = \mathbf{A},\ \mathbf{DB} = \mathbf{C}. \end{aligned} \tag{3}$$

Now, Problem 3 matches robust PCA, which can be solved using augmented Lagrange multiplier (ALM) method [15]. With ALM, Problem 3 is reformulated as

$$\begin{aligned}\min_{\mathbf{B},\mathbf{E},\mathbf{N},\mathbf{A},\mathbf{C}} \; & \|\mathbf{A}\|_* + \lambda_1\|\mathbf{E}\|_1 + \frac{\lambda_2}{2}\|\mathbf{N}\|_F^2 \; + \lambda_3\|\mathbf{W}_b \circ \mathbf{C}\|_1 + \\ & \frac{\mu}{2}\|\mathbf{F}-\mathbf{B}-\mathbf{E}-\mathbf{N}\|_F^2 + \frac{\mu}{2}\|\mathbf{B}-\mathbf{A}\|_F^2 + \\ & \frac{\mu}{2}\|\mathbf{D}\mathbf{B}-\mathbf{C}\|_F^2 \; + \langle \mathbf{Y}_1, \mathbf{F}-\mathbf{B}-\mathbf{E}-\mathbf{N}\rangle + \\ & \langle \mathbf{Y}_2, \mathbf{B}-\mathbf{A}\rangle + \langle \mathbf{Y}_3, \mathbf{D}\mathbf{B}-\mathbf{C}\rangle,\end{aligned} \tag{4}$$

where $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$ are the Lagrange multipliers, μ is a penalty parameter, and $\langle \cdot, \cdot \rangle$ denote the sum of product of corresponding matrix elements.

An important operator used in various solutions of RPCA is the *soft thresholding* operator [4, 15], which is applied to each matrix element individually:

$$S_\varepsilon(x) = \begin{cases} x-\varepsilon, \text{ if } x > \varepsilon, \\ x+\varepsilon, \text{ if } x < -\varepsilon, \\ 0, \qquad \text{otherwise.} \end{cases} \tag{5}$$

With this operator, [3] show that, for matrix $\mathbf{M}$ with SVD $\mathbf{U}\mathbf{S}\mathbf{V}^\top$, the minimizations of nuclear norm and l_1-norm are given by

$$\mathbf{U}\, S_\varepsilon(\mathbf{S})\mathbf{V}^\top = \arg\min_{\mathbf{X}} \varepsilon\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{M}-\mathbf{X}\|_F^2, \tag{6}$$

$$S_\varepsilon(\mathbf{M}) = \arg\min_{\mathbf{X}} \varepsilon\|\mathbf{X}\|_1 + \frac{1}{2}\|\mathbf{M}-\mathbf{X}\|_F^2. \tag{7}$$

Applying Eq. 6 and 7 to the optimization of $\mathbf{A}$, $\mathbf{E}$, and $\mathbf{C}$ in Problem 4 give

$$\mathbf{A} = \mathbf{U} S_{1/\mu}(\mathbf{S})\mathbf{V}^\top, \tag{8}$$

$$\mathbf{E} = S_{\lambda_1/\mu}\left(\mathbf{F}-\mathbf{B}-\mathbf{N}+\mathbf{Y}_1/\mu\right), \tag{9}$$

$$\mathbf{C} = S_{\lambda_3\mathbf{W}_b/\mu}\left(\mathbf{D}\mathbf{B}+\mathbf{Y}_3/\mu\right), \tag{10}$$

where $\mathbf{U}\mathbf{S}\mathbf{V}^\top$ is the SVD of $\mathbf{B}+\mathbf{Y}_2/\mu$. In our problem, the ranks of the static background and reflection are known to be 1. Therefore, Eq. 8 is equivalent to

$$\mathbf{A} = \mathbf{U}\mathbf{S}_1\mathbf{V}^\top, \tag{11}$$

where $\mathbf{S}_1$ is a diagonal matrix that contains only the first singular value of $\mathbf{S}$.

The optimization of $\mathbf{N}$, with other variables fixed, involves only Frobenius norm, and can be derived directly as

$$\mathbf{N} = \frac{\mu}{\lambda_2+\mu}\left(\mathbf{F}-\mathbf{B}-\mathbf{E}+\frac{1}{\mu}\mathbf{Y}_1\right). \tag{12}$$

Now, with $\mathbf{A}$, $\mathbf{E}$, $\mathbf{C}$, and $\mathbf{N}$ fixed, Problem 4 reduces to

$$\begin{aligned}\min_{\mathbf{B}} \; & \frac{\mu}{2}\|\mathbf{B}-(\mathbf{F}-\mathbf{E}-\mathbf{N}+\frac{1}{\mu}\mathbf{Y}_1)\|_F^2 + \\ & \frac{\mu}{2}\|\mathbf{B}-(\mathbf{A}-\frac{1}{\mu}\mathbf{Y}_2)\|_F^2 \; + \frac{\mu}{2}\|\mathbf{D}\mathbf{B}-(\mathbf{C}-\frac{1}{\mu}\mathbf{Y}_3)\|_F^2,\end{aligned}$$

which can be rearranged as

$$\min_{\mathbf{B}} \frac{\mu}{2} \left(\mathbf{B}^\top (\mathbf{D}^\top \mathbf{D} + 2\mathbf{I}) \mathbf{B} - \mathbf{T}^\top \mathbf{B} \right), \tag{13}$$

where

$$\mathbf{T} = \mathbf{F} - \mathbf{E} - \mathbf{N} + \frac{1}{\mu}\mathbf{Y}_1 + \mathbf{A} - \frac{1}{\mu}\mathbf{Y}_2 + \mathbf{D}^\top \left(\mathbf{C} - \frac{1}{\mu}\mathbf{Y}_3 \right).$$

Problem 13 is a quadratic optimization problem whose solution is given by [5]

$$(\mathbf{D}^\top \mathbf{D} + 2\mathbf{I})\mathbf{B} = \mathbf{T}. \tag{14}$$

Eq. 14 can be efficiently solved using fast Fourier transform (FFT) [5] as follows: Let $\mathbf{b}_i$ and $\mathbf{t}_i$ denote the 2D forms of column i of $\mathbf{B}$ and $\mathbf{T}$ respectively. Then,

$$\mathbf{b}_i = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(\mathbf{t}_i)}{\overline{\mathcal{F}(\mathbf{D})} \circ \mathcal{F}(\mathbf{D}) + \mathbf{2}} \right), \tag{15}$$

where $\mathcal{F}$, $\mathcal{F}^{-1}$, and $\overline{\mathcal{F}}$ are the 2D FFT, 2D inverse FFT and the complex conjugate of 2D FFT, respectively, and $\mathbf{2}$ is a matrix whose elements are all 2. The division in Eq. 15 is performed element-wise.

Finally, the Lagrange multipliers and parameter μ are updated according to ALM method [15] as follows:

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{Y}_1 + \mu(\mathbf{F} - \mathbf{B} - \mathbf{E} - \mathbf{N}), \\ \mathbf{Y}_2 &= \mathbf{Y}_2 + \mu(\mathbf{B} - \mathbf{A}), \\ \mathbf{Y}_3 &= \mathbf{Y}_3 + \mu(\mathbf{D}\mathbf{B} - \mathbf{C}), \\ \mu &= \rho\mu, \text{ for } \rho > 1. \end{aligned} \tag{16}$$

The variables discussed in this section are updated iteratively according to ALM until convergence (Algorithm 2).

2.4 Updating of Reflection Transformations

In our alternating optimization scheme, updating of transformations T_r and T_r' of reflection is performed before updating of the aligned reflection $\mathbf{R}$. After the aligned background $\mathbf{B}$ is obtained (Section 2.3), the unaligned reflection is computed as $\mathbf{R}' = \mathbf{F}' - T_b'(\mathbf{B})$. Among the reflection images $\mathbf{r}_i'$ in $\mathbf{R}'$, the one that corresponds to the neutral viewing angle, denoted as $\mathbf{r}_0'$, is selected as reference. All other reflection images $\mathbf{r}_i'$ are aligned to $\mathbf{r}_0'$. That is, the transformation of $\mathbf{r}_0'$ is identity function, giving $\mathbf{r}_0' = \mathbf{r}_0$.

The transformation between $\mathbf{r}_i'$ and $\mathbf{r}_0' = \mathbf{r}_0$ is computed as follows. First, matching pairs of feature points are extracted from $\mathbf{r}_i'$ and $\mathbf{r}_0$ based on scale-invariant feature transform (SIFT) [17] and speeded up robust features (SURF) [2]. The matching pairs may contain undesirable features that belong to the background instead of the reflection. Consider a matching pair (p_i, p_0). If their positions $T_b(p_i)$ and $T_b(p_0)$ after aligning to the background $\mathbf{B}$ are close to each other, then they belong to the background instead of the reflection. So, they are removed from matching pairs for computing T_r and T_r'.

As the reflection scene is non-planar, a single homography may not be sufficient to model the transformation. So, RANSAC is applied to find the largest set of consistent matching pairs that fits a homography. Then, the consistent set is removed and

RANSAC is applied to find the next largest set of consistent matching pairs. This procedure is repeated until there are too few remaining matching pairs. If the number of consistent set is one, then a pair of homographies can be computed from the consistent set as T_r and T'_r. Otherwise, thin-plate spline is used to derive T_r and T'_r from the union of the consistent sets.

This procedure is repeated for each $\mathbf{r}'_i$. The transformations for different $\mathbf{r}'_i$ are different. Initially, the transformations computed are between $\mathbf{r}'_i$ and the reference $\mathbf{r}_0$. As the algorithm iterates, the rank-1 matrix $\mathbf{R}$ converges such that $\mathbf{r}_i = \mathbf{r}_0$ for all i. Then, desired transformations between $\mathbf{r}'_i$ and $\mathbf{r}_i$ are obtained.

2.5 Updating of Background Weights

There are three kinds of edges in the images: (1) edges that belong to the background, (2) edges that belong to the reflection, and (3) phantom edges that result from separating non-edge pixels into background and reflection. To discriminate between these edges, three gradient fields are computed from the aligned images, namely $\mathbf{G}_f = \mathbf{D}T_b(\mathbf{F}')$, $\mathbf{G}_r = \mathbf{D}T_b(T'_r(\mathbf{R}))$, and $\mathbf{G}_b = \mathbf{G}_f - \mathbf{G}_r$. Note that $T_b(\mathbf{F}')$ and $T_b(T'_r(\mathbf{R}))$ make up $\mathbf{F}$ that is used for optimization of background (Section 2.3). Salient edges are extracted by thresholding the gradient fields into corresponding binary matrices $\mathbf{H}_f$, $\mathbf{H}_b$, and $\mathbf{H}_r$ at threshold τ. Then, element $[j,k]$ of background weight $\mathbf{W}_b$, denoted as $\mathbf{W}_b[j,k]$, is updated as follows:

$$\mathbf{W}_b[j,k] = \begin{cases} 0, \text{ if majority of } \mathbf{H}_f[j,l], \text{for } l = 1,\ldots,n, \text{ satisfies} \\ \quad \mathbf{H}_f[j,l] = 1 \text{ and } |\mathbf{G}_b[j,l]| > \alpha|\mathbf{G}_r[j,l]|, \\ 5, \text{ if } \mathbf{H}_f[j,k] = 1 \text{ and } |\mathbf{G}_r[j,k]| > \alpha|\mathbf{G}_b[j,k]|, \\ 5, \text{ if } \mathbf{H}_f[j,k] = 0 \text{ and } (|\mathbf{G}_b[j,k]| > \tau \text{ or } |\mathbf{G}_r[j,k]| > \tau), \\ 1, \text{ otherwise.} \end{cases} \tag{17}$$

The first three conditions correspond, respectively, to background edges, reflection edges, and phantom edges. In other words, background edges are associated with zero weight whereas non-background edges are associated with large weights. The first condition fills the whole row j of $\mathbf{W}_b$ with 0 if the majority of the elements in row j of $\mathbf{H}_f$ satisfies the condition. This condition overrides the 2nd and 3rd conditions. This method facilitates the convergence of $\mathbf{B}$ into a rank-1 matrix. The last condition provides default weight of 1 for the other pixels. In the current implementation, $\tau = 0.05$ and $\alpha = 1.2$.

2.6 Initialization

The initialization of our algorithm CLORD is similar to the min-max alternation method of [23] for reflection separation. Among the superimposed images $\mathbf{f}'_i$ in $\mathbf{F}'$, the one at the neutral viewing angle, denoted as $\mathbf{f}'_0$, is selected as the reference. Since the image intensity is dominated by background intensity, we can apply the transformation update algorithm described in Section 2.4 on $\mathbf{f}'_i$ to estimate the transformations T_b and T'_b of the background. Next, an initial estimate of the background image $\mathbf{b}$ is obtained by finding the row-minima of $T_b(\mathbf{F}')$, which is inserted into each column of $\mathbf{B}$. Next, $\mathbf{R}'$ is computed as $\mathbf{F}' - T'_b(\mathbf{B})$. Then, the transformation update algorithm is applied on $\mathbf{r}'_i$ to estimate the transformations T_r and T'_r of the reflection. Finally, an initial estimate of the reflection image $\mathbf{r}$ is obtained by finding the row-maxima of $T_r(\mathbf{R}')$, which is inserted into each column of $\mathbf{R}$.

2.7 Summary

Our CLORD method handles background $\mathbf{B}$ and reflection $\mathbf{R}$ in a uniform manner. So, the optimization of $\mathbf{R}$, updating of background transformations T_b and T_b', and updating of reflection weights $\mathbf{W}_r$ are achieved by applying the same algorithms described in the preceding sections, with $\mathbf{B}$ and $\mathbf{R}$ swapped. The complete algorithm is summarized in Algorithm 1, and the optimization of rank-1 matrix $\mathbf{B}$ (as well as $\mathbf{R}$, with $\mathbf{B}$ and $\mathbf{R}$ swapped) is summarized in Algorithm 2. In the current implementation, λ_1, λ_2, λ_3 are set to 50λ, 2000λ, and 10λ respectively, where $\lambda = 1/\sqrt{\max(m,n)}$ [4].

Algorithm 1: Coupled Low-Rank Decomposition (CLORD)

Input: $\mathbf{F}'$
Initialize $\mathbf{B}$, $\mathbf{R}$, T_b, T_b', T_r, T_r' (Section 2.6).
repeat
  Update the weight matrix $\mathbf{W}_b$ (Section 2.5).
  Optimize background $\mathbf{B}$ (Algorithm 2, Section 2.3).
  Estimate the transformations of reflection T_r and T_r' (Section 2.4).
  Update the weight matrix $\mathbf{W}_r$ (Section 2.5).
  Optimize reflection $\mathbf{R}$ (Algorithm 2, Section 2.3).
  Estimate the transformations of background T_b and T_b' (Section 2.4).
until *convergence*;
Output: $\mathbf{B}$, $\mathbf{R}$.

Algorithm 2: Updating of Rank-1 Matrix

Input: $\mathbf{F}$
Initialize $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$, $\mu > 0$, $\rho > 1$.
Initialize $\mathbf{B}, \mathbf{A}, \mathbf{E}, \mathbf{C}, \mathbf{N}$ to $\mathbf{0}$.
repeat
  $\mathbf{U}, \mathbf{S}, \mathbf{V} = \text{SVD}(\mathbf{B} + \mathbf{Y}_2/\mu)$.
  $\mathbf{A} = \mathbf{U}\mathbf{S}_1\mathbf{V}^\top$.
  $\mathbf{E} = S_{\lambda_1/\mu}(\mathbf{F} - \mathbf{B} - \mathbf{N} + \mathbf{Y}_1/\mu)$.
  $\mathbf{C} = S_{\lambda_3 \mathbf{w}_b/\mu}(\mathbf{D}\mathbf{B} + \mathbf{Y}_3/\mu)$.
  $\mathbf{N} = \frac{\mu}{\lambda_2+\mu}(\mathbf{F} - \mathbf{B} - \mathbf{E} + \mathbf{Y}_1/\mu)$.
  Update $\mathbf{B}$ according to Eq. 15.
  $\mathbf{Y}_1 = \mathbf{Y}_1 + \mu(\mathbf{F} - \mathbf{B} - \mathbf{E} - \mathbf{N})$.
  $\mathbf{Y}_2 = \mathbf{Y}_2 + \mu(\mathbf{B} - \mathbf{A})$.
  $\mathbf{Y}_3 = \mathbf{Y}_3 + \mu(\mathbf{D}\mathbf{B} - \mathbf{C})$.
  $\mu = \rho\mu$.
until *convergence*;
Output: $\mathbf{B}$

3 Experiments and Discussions

This section compares the performance of CLORD and those of recent existing methods denoted as LI [14], GUO [8], and XUE [24]. As the program for XUE is not publicly available, we compare it only with the results published in [24] (Section 3.1, 3.2). To handle color images, the methods are applied to each of the R, G and B channels and their outputs are combined into the final results. Test programs were implemented in Matlab, and ran on a PC with Intel Core 3.5GHZ CPU and 32GB RAM.

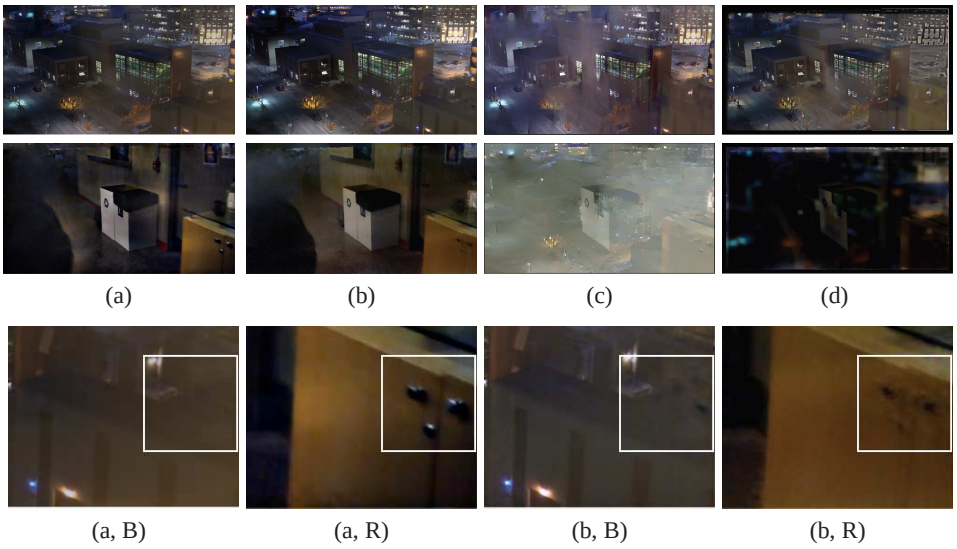


Fig. 1. Test results of night scene images with planar background and reflection. (Row 1) Recovered background, (row 2) separated reflection, (row 3) zoom-in views of bottom-right corner of the images. White boxes show that XUE's background has remnants of reflection. (a) CLORD, (b) XUE, (c) LI, (d) GUO. (B) Background, (R) reflection.

3.1 Planar Background and Reflection Scenes

This test evaluates the methods' base-case performance on images of planar background and reflection scenes. This scenario is valid when all objects are far away from the camera. 5 images of a night scene captured through a window, and 5 images of indoor scene with glass reflection were used for the test. These images were obtained from [24] and ground truth was not available.

For the night scene, Figure 1 shows that the results of CLORD and XUE are visually accurate whereas GUO and LI cannot separate the background and the reflection well. This could be because GUO and LI explicitly optimize the background but not the reflection, which implies that better reflection recovery can improve background recovery. There is a subtle difference between the results of CLORD and XUE (Fig. 1, row 3). The background recovered by XUE has some remnants of reflection. On the other hand, the background recovered by CLORD does not have reflection, and its recovered reflection is much more complete than that of XUE. This difference could be due to XUE's non-locally linear flow filed.

The indoor results of CLORD and XUE are shown in figure 2(b) and 2(c). Consistent with the previous test, XUE's background has more remnants of reflection compared to that of CLORD. Figure 2(d) and 2(e) show that input images transformed by the computed T_b align very well to the reference background. Figure 2(f) visualizes the weight W_b of the reference image in figure 2(a). It shows that the edge of background and reflection objects are well separated. In this test, CLORD takes 8 minutes to process 5 images of size 1152×648 whereas XUE takes about 20 minutes.

(a) (b) (c) (d) (e) (f)

Fig. 2. Test results of indoor scene images with planar background and reflection. (a) Reference image, (b) CLORD, (c) XUE, (d, e) input images transformed by the computed T_b align very well to the reference background in (a), (f) W_b associated with (a). As shown in (b) and (c), XUE's background has more remnants of relection compared to that of CLORD, their recovered reflections are similar and are omitted. In (f), background edges (dark) and reflection edges (light) are well separated.

Table 1. Normalized cross correlation of various methods tested on images with non-planar background and reflection scenes.

method	CLORD	XUE	LI	GUO
background	**0.94**	0.90	0.79	0.77
reflection	**0.83**	0.75	0.61	0.69

3.2 Non-Planar Background and Reflection Scenes

This test evaluates the methods' performance on images of non-planar background and reflection scenes. This scenario is valid when the objects are located at different depths near the camera. 5 images from [24] captured at different viewing angles were used, and ground truths of the background and reflection were available. The degree of match between the recovered images and the ground truths were measured as normalized cross correlation (NCC).

Figure 3 shows that the results of CLORD and XUE are visually accurate. On the other hand, GUO and LI cannot separate the background and the reflection well because they do not explicitly optimize the reflection. In addition, GUO is not designed to handle non-planar scenes. In comparison, CLORD's recovered background is more accurate, and XUE's background has some distortions around the head of the toy frog. Table 1 confirms that CLORD is more accurate than the other methods for non-planar scenes.

3.3 Scene Distance

This experiment investigates the effect of the distances of background objects and reflection objects from the reflecting glass on the methods' performance. A background object was placed behind a piece of glass, and a reflection object was placed in front. The ratio d_r of the distances of the reflection object and the background object from the glass was varied from $4/8$ to $8/8 = 1$ in steps of $1/8$. For each distance ratio d_r, ground truths and 5 superimposed images were captured at different viewing angles.

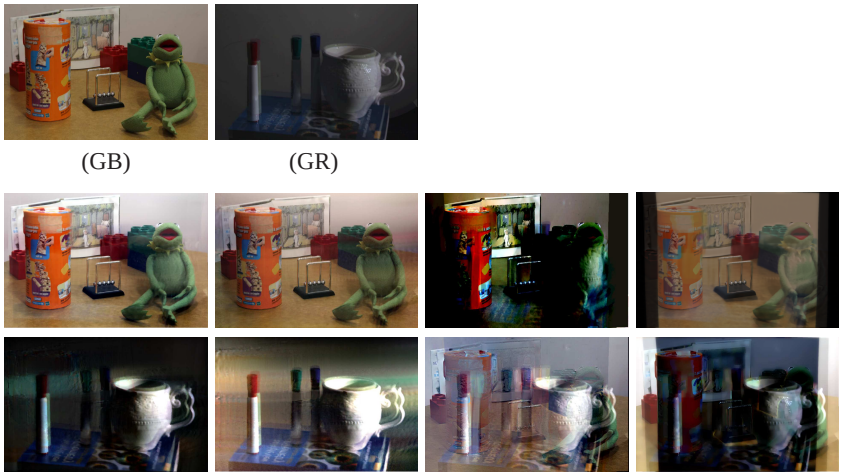


Fig. 3. Test results on images of non-planar background and reflection scenes. (GB) Ground truth background, (GR) ground truth reflection. (Row 2) Recovered background, (row 3) separated reflection. (a) CLORD, (b) XUE, (c) LI, (d) GUO.

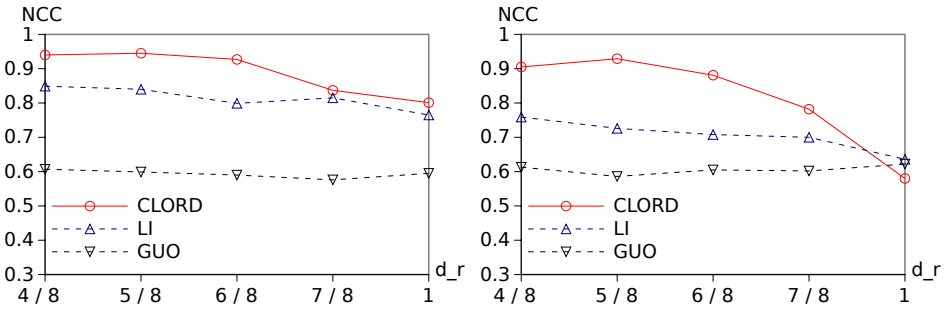
The methods were tested, and NCC between ground truths and recovered images were measured.

Figure 4 shows that CLORD is more accurate than LI and GUO. CLORD's background NCC exceeds 0.9 for $d_r \leq 6/8$. On the other hand, LI's background NCC is less than 0.9, and GUO's background NCC hovers around 0.6. The accuracies of CLORD and LI decrease gradually with increasing d_r as expected because the background scene and virtual scene become more coplanar. The accuracy of GUO remains low for all d_r. The methods' reflection NCC shows a similar trend as their background NCC.

3.4 Degree of Non-Planarity

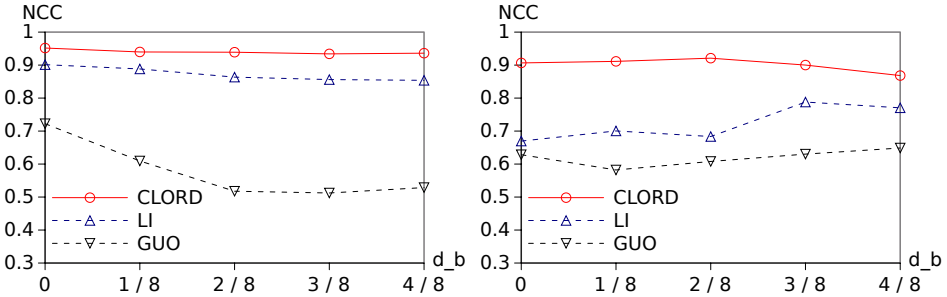
This experiment investigates the effect of the degree of non-planarity on the methods' performance. Two background objects, were placed behind a piece of glass, and a reflection object was placed in front. The ratio of the distance of the reflection object and the nearer background object from the glass was set at $1/2$. The further background object was placed behind the nearer object, and its distance ratio d_b from the nearer object varied from 0 to $4/8$ in steps of $1/8$. For each distance ratio d_b, ground truths and 5 superimposed images were captured at different viewing angles. The methods were tested, and NCC between ground truths and recovered images were measured.

Figure 5 shows that CLORD's background NCC exceeds 0.9 for all d_b, whereas LI's background NCC is less than 0.9. GUO's background NCC is slightly above 0.7 at $d_b = 0$ when the background scene is planar, and it decreases with increasing d_b, which makes the background scene more and more non-planar. CLORD's reflection NCC is around 0.9 for all d_b. LI's reflection NCC is slightly below 0.7 at $d_b = 0$, and it increases with increasing d, suggesting that LI can identify reflection more accurately for non-



(a) Background (b) Reflection

Fig. 4. Normalized cross correlation (NCC) of various methods at varying distance ratio d_r of reflection with respect to background. Smaller d_r means greater depth disparity between background and reflection objects.



(a) Background (b) Reflection

Fig. 5. Normalized cross correlation (NCC) of various methods at varying distance ratio d_b of the further background object from the nearer background object. Larger d_b means greater non-planarity.

planar scenes. GUO's reflection NCC is slightly above 0.6. The test results show that CLORD is more accurate than LI and GUO for the non-planar scenes.

4 Conclusion

This paper presented a multiple-view method for separating reflection and background in unaligned images by coupled low-rank decomposition (CLORD). It is based on the method of robust PCA. It models the background and reflection as rank-1 matrices in a uniform manner, which are decomposed according to different transformations that align the background images and the reflection images separately. Comprehensive test results show that CLORD is more accurate and robust than recent related methods, especially for images with non-planar scenes and global reflection.

References

1. A. Agrawal, R. Raskar, R. K. Nayar, and Y. Z. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Trans. Graphics*, 24(3), 2005.

2. H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 2008.
3. J. F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optimization*, 20(4), 2010.
4. E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3), 2011.
5. S. Chaudhuri, R. Velmurugan, and R. Rameshan. *Blind Image Deconvolution: Methods and Convergence*. Springer, 2014.
6. H. Farid and E. H. Adelson. Separating reflections from images by use of independent component analysis. *J. Optical Society of America A*, 16(9), 1999.
7. K. Gai, Z. W. Shi, and C. S. Zhang. Blindly separating mixtures of multiple layers with spatial shifts. In *Proc. CVPR*, 2008.
8. X. Guo, X. Cao, and Y. Ma. Robust separation of reflection from multiple images. In *Proc. CVPR*, 2014.
9. X. Jiang and J. Lai. Sparse and dense hybrid representation via dictionary decomposition for face recognition. *IEEE Trans. PAMI*, 37(5), May 2015.
10. N. Kong, Y. Tai, and J. S. Shin. A physically-based approach to reflection separation: From physical modeling to constrained optimization. *IEEE Trans. PAMI*, 36(2), 2014.
11. W. K. Leow, Y. Cheng, L. Zhang, T. Sim, and L. Foo. Background recovery by fixed-rank robust principal component analysis. In *Proc. CAIP*, 2013.
12. A. Levin and Y. Weiss. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Trans. PAMI*, 29(9), 2007.
13. A. Levin, A. Zomet, and Y. Weiss. Separating reflections from a single image using local features. In *Proc. CVPR*, 2004.
14. Y. Li and M. S. Brown. Exploiting reflection change for automatic reflection removal. In *Proc. ICCV*, 2013.
15. Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical report, UIUC, 2009.
16. Z. Lou and T. Gevers. Image alignment by piecewise planar region matching. *IEEE Trans. Multimedia*, 16(7), 2014.
17. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Computer Vision*, 60(2), 2004.
18. T. H. Oh, J. Y. Lee, Y. W. Tai, and I. S. Kweon. Robust high dynamic range imaging by rank minimization. *IEEE Trans. PAMI*, 37(6), 2015.
19. T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE Trans. PAMI*, 38(4), 2016.
20. Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. PAMI*, 34(11), 2012.
21. Y. Y. Schechner, N. Kiryati, and R. Basri. Separation of transparent layers using focus. *Int. J. Computer Vision*, 39(1), 2000.
22. Y. Y. Schechner, J. Shamir, and N. Kiryati. Polarization-based decorrelation of transparent layers: The inclination angle of an invisible surface. In *Proc. ICCV*, 1999.
23. R. Szeliski, S. Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. In *Proc. CVPR*, 2000.
24. T. F. Xue, M. Rubinstein, C. Liu, and W. T. Freeman. A computational approach for obstruction-free photography. *ACM Trans. Graphics*, 34(4), 2015.
25. J. L. Yang, H. D. Li, Y. C. Dai, and R. T. Tan. Robust optical flow estimation of double-layer images under transparency or reflection. In *Proc. CVPR*, 2016.