

# TIME-RESOLVED IMAGE DEMIXING

Ayush Bhandari<sup>†</sup>, Aurélien Bourquard<sup>‡</sup>, Shahram Izadi<sup>○</sup>, and Ramesh Raskar<sup>†</sup>

{<sup>†</sup>Media Laboratory, <sup>‡</sup>RLE}, Massachusetts Institute of Technology, Cambridge, USA.

<sup>○</sup>Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA.

Email: ayush@MIT.edu • aurelien@MIT.edu • shahrami@microsoft.com • raskar@MIT.edu

## ABSTRACT

When multiple light paths combine at a given location on an image sensor, an image mixture is created. Demixing or recovering the original constituent components in such cases is a highly ill-posed problem. A number of elegant solutions have thus been developed in the literature, relying on measurement diversity such as polarization, shift, motion, or scene features. In this paper, we approach the image-mixing problem as a time-resolved phenomenon—if every photon arriving at the sensor could be time-stamped, the demixing problem would then amount to separating transient events in time. Based on this idea, we first show that, while acquiring measurements is prohibitive and challenging in the time domain, this task is surprisingly straightforward in the frequency domain. We then establish a link between frequency-domain measurements and consumer time-of-flight (ToF) imaging. Finally, we propose a demixing algorithm, relying only on magnitude information of the ToF sensor. We show that our problem is closely tied to the topic of phase retrieval and that for  $K$ -image mixture,  $(K^2 - K)/2 + 1$  magnitude-only ToF measurements suffice to demix images exactly in noiseless settings. Our developments are corroborated with experiments on synthetic and ToF data acquired using the Microsoft Kinect sensor.

**Index Terms**— Image demixing, phase retrieval, spectral estimation, time-of-flight cameras, time-resolved imaging.

## 1. INTRODUCTION

Imaging modalities work on the premise that there is one-to-one mapping between the scene and the pixels. If this were not the case, it would be almost impossible to make sense of the photographs that we capture on a daily basis. Meanwhile, the illustration in Fig. 1 is a characteristic counter-example. This is the main theme of this work:

Can we recover images when multiple optical paths mix at the sensor or when there exists a many-to-one mapping between the scene and the pixel?

Such settings have been addressed in a number of papers during the last decade [1–5]. Since multiple light paths superimpose at the sensor, we are interested in **demixing** the constituent components so that the image of interest can be recovered from the image mixture. Almost all of the papers that deal with this issue of demixing sources study this problem in context of **Blind Source Separation** (or BSS). The key idea is to exploit image priors to regularize the inverse problem at hand.

In this paper, we take a first approach to repurpose Time-of-Flight (ToF) cameras to demix light paths in everyday scenes. Unlike BSS methods that assume image priors, our work explores the idea of exploiting depth diversity in the scene.

Let us assume for a moment that we can label each photon with a time stamp indicating the time at which it reached the sensor. In such a setting, it would not be too outrageous to conclude that the problem of “demixing” the image boils down to the problem of demixing



**Fig. 1:** Photograph taken at a train station in Bern, Switzerland that illustrates the mixing of light paths and hence image superposition.

events in time. We explain this with the help of a schematic in Fig. 2. In Fig. 2 (a), we mimic the image-formation process of a conventional camera where two optical paths combine at the sensor. In Fig. 2 (b), we discuss the same setting assuming that we have time-resolved measurements. With reference to Fig. 2, we note that, in theory, the two images— $\Gamma_1(x, y)$  and  $\Gamma_2(x, y)$ —mix at distinct time instants  $t = t_1$  and  $t = t_2$  together with mixing coefficients  $\alpha_1(x, y)$  and  $\alpha_2(x, y)$ , respectively, thus producing the time-resolved or transient image

$$m(x, y, t) = (\alpha_1 \Gamma_1)(x, y) \delta(t - t_1) + (\alpha_2 \Gamma_2)(x, y) \delta(t - t_2). \quad (1)$$

In (1), the symbol  $\delta$  denotes the Dirac impulse. For the sake of simplicity, we define  $(\alpha \Gamma)(x, y) \stackrel{\text{def}}{=} \alpha(x, y) \Gamma(x, y)$ . The time-resolved image (TRI)  $m(x_0, y_0, t)$  is shown in Fig. 2 (b). Given the exposure time  $\tau \gg t_2$ , a conventional camera acquires the image

**Conventional Camera**

$$\int_0^\tau m(x, y, t) dt = \underbrace{(\alpha_1 \Gamma_1)(x, y) + (\alpha_2 \Gamma_2)(x, y)}_{m(x, y)}, \quad (2)$$

where the dependency on  $t$  in  $m(x, y)$  is dropped for the rest of this paper. This should be understood as  $m(x, y) = \int_0^\tau m(x, y, t) dt$ . When extended to the case of  $K$  semi-reflective surfaces,

$$m(x, y, t) = \sum_{k=0}^{K-1} (\alpha_k \Gamma_k)(x, y) \delta(t - t_k), \quad (3)$$

and, consequently, by marginalizing time variable, a camera records

$$m(x, y) \stackrel{(2)}{=} \sum_{k \in \{k | \tau \geq t_k\}} (\alpha_k \Gamma_k)(x, y). \quad (4)$$

Now, even in the most elementary case  $K = 2$ , the problem of recovering  $\Gamma_{1,2}$  is severely ill-posed.

## 2. TOF IMAGE FORMATION MODEL

ToF sensors are active homodyne devices equipped with lock-in detection functionality [6]. A ToF camera probes the scene with a sinusoidal function of the form  $p(t) = 1 + p_0 \cos(\omega_0 t)$ . At a given pixel  $(x_0, y_0)$  and at a distance  $z_0$  from the sensor, the probing function interacts with the scene and results in the reflected signal  $r(t) = \Gamma_0 p(t - t_0)$ ,  $t_0 = 2z_0/c$ , where  $c$  denotes the speed of light. The sensor then cross-correlates  $r(t)$  and  $p(t)$  to obtain [7, 8]

$$C_{\omega_0}(t) = \Gamma_0 \left( 1 + \frac{p_0^2}{2} \cos(\omega_0(t + t_0)) \right). \quad (5)$$

Knowing  $\{\Gamma_0, t_0\}$  amounts to knowing the whole 3D scene. This is accomplished via the “four–bucket principle” [6–8]. More precisely, given  $c_k = C_{\omega_0}(\pi k / 2\omega_0)$ ,  $k \in \mathbb{Z}$ , we define a complex number,  $Z_{\omega_0} = (c_0 - c_2) + j(c_3 - c_1)$ . One can then define the estimates

ToF Four Bucket Principle

$$\tilde{\Gamma}_0 = |Z_{\omega_0}| / p_0^2, \quad \tilde{t}_0 = \angle Z_{\omega_0} / \omega_0, \quad \hat{m}(\omega_0) = \tilde{\Gamma}_0 e^{-j\omega_0 \tilde{t}_0} \quad (6)$$

where, having estimated  $\{\tilde{\Gamma}_0, \tilde{t}_0\}$ , we constitute the complex number  $\hat{m}(\omega) = \tilde{\Gamma}_0 e^{-j\omega_0 \tilde{t}_0}$  [7, 8]. With  $t_0 = 2z_0/c$ , it follows from the above that, for each pixel, the information corresponds to [8, 9]:

$$\hat{m}(\omega) = \underbrace{\sum_{k=0}^{K-1} \Gamma_k e^{-j\omega t_k}}_{\text{Fourier Time-Resolved Image (FTRI)}} \xrightarrow{\text{Fourier}} m(t) = \underbrace{\sum_{k=0}^{K-1} \Gamma_k \delta(t - t_k)}_{\text{Time-Resolved Image}}. \quad (7)$$

Now, in contrast to prior work linked with AMCW–ToF [7, 9–11] and references therein, our assumption in this paper is that only the *amplitude* information of the measurements can be accessed.

### 2.1. Phase Retrieval for ToF Imaging

The above definitions imply that phase information encodes the parameters  $t_k$ . Now, the quantities that are of interest to us consist in the time-resolved-image intensities. In the case  $K = 1$ , we have

$$|\hat{m}(\omega_0)|^2 = |\Gamma_0|^2, \quad (\text{measurement}) \quad (8)$$

which is the intensity corresponding to the time-resolved image  $m(t) = \Gamma_0 \delta(t - t_0)$ . Such measurements are computed using the auto-correlation function,

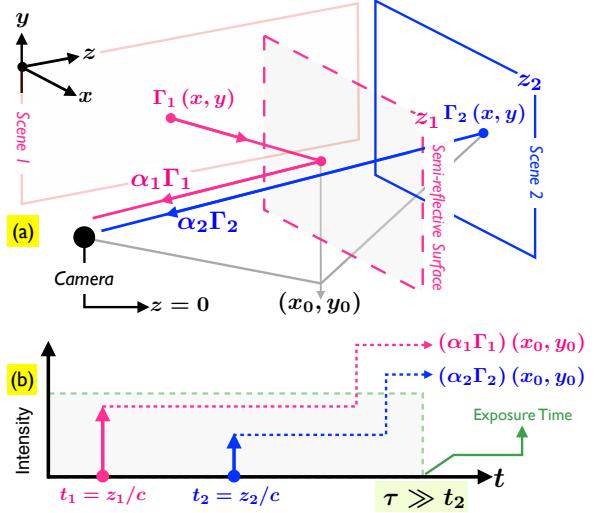
$$A_{\omega_0}(t) = (C_{\omega_0} * \bar{C}_{\omega_0})(t) = \frac{\Gamma_0^2}{8} (8 + p_0^4 \cos(\omega_0 t)). \quad (9)$$

Letting  $a_k = A_{\omega_0}(\pi k / 2\omega_0)$ , one can now estimate  $|\tilde{\Gamma}_0|^2$  using two measurements  $|\hat{m}(\omega_0)|^2 = |\tilde{\Gamma}_0|^2 = 4(a_0 - a_2) / p_0^4$ . This procedure bypasses phase computations producing a single real-valued measurement per ToF exposure unlike the usual two-value case (6).

When  $K$  light paths meet at the sensor (cf. Fig. 2), following (9), for a given modulation frequency  $\omega$ , the ToF lock-in sensor records,

$$\begin{aligned} |\hat{m}(\omega)|^2 &= \left| \sum_{k=0}^{K-1} \Gamma_k e^{-j\omega t_k} \right|^2 \quad (\text{from (6) and (7)}) \\ &= \sum_{k=0}^{K-1} |\Gamma_k|^2 + 2 \sum_{k=0}^{K-1} \sum_{l=k+1}^{K-1} |\Gamma_k| |\Gamma_l| \cos(\omega t_{kl} + \angle \Gamma_{kl}), \end{aligned} \quad (10)$$

where  $t_{kl} = t_k - t_l$  and  $\angle \Gamma_{kl} = \angle \Gamma_k - \angle \Gamma_l$ . Knowing  $\hat{m}(\omega)$  is fundamentally different from knowing  $|\hat{m}(\omega)|^2$  because the former encodes both phase and amplitude. As shown in [7, 8], given  $\hat{m}(\omega)$ ,



**Fig. 2:** Conventional and time–resolved imaging. (a) Image-mixing process: The scene comprises of a semi–reflecting mirror at  $z = z_1$  followed by the profile to be photographed at  $z = z_2$ . At given pixel  $(x_0, y_0)$  inside the camera—itself located at  $(x_0, y_0, 0)$ —two additive light paths with intensities  $(\alpha_1 \Gamma_1)(x_0, y_0)$  and  $(\alpha_2 \Gamma_2)(x_0, y_0)$  are observed,  $\alpha_1, \alpha_2$  being mixing coefficients associated with the reflecting surface. (b) Time–resolved/transient image  $m(x_0, y_0, t)$ : If we could determine the time at which photons impact the camera/sensor, the intensity  $(\alpha_1 \Gamma_1)(x_0, y_0)$  would correspond to photons that take time  $t = z_1/c$  where  $c$  is the speed of light. Similarly,  $(\alpha_2 \Gamma_2)(x_0, y_0)$  would be attributed to the time instant  $t = z_2/c$ . Hence, demixing involves identification of photons linked with instants  $z_1/c$  and  $z_2/c$ . Such transient information is not available with conventional cameras because of the too long exposure time  $\tau \gg t_2$ .

one can discern  $\{\Gamma_k, t_k\}_{k=0}^{K-1}$  unlike in the phase-less case. Our goal is to demix images for which we only require to estimate  $\{\Gamma_k\}_{k=0}^{K-1}$ . We will show that one can indeed recover constituent images  $\{\Gamma_k\}_{k=0}^{K-1}$  with phase-less  $|\hat{m}(n\omega_0)|^2$ .

## 3. RECONSTRUCTION VIA PHASE RETRIEVAL

Let us re-arrange (10) based on Euler’s identity to obtain a variant of the *Line Spectrum Estimation* or the LSE problem [12]:

$$y_n = |\hat{m}(n\omega_0)|^2 \stackrel{(10)}{=} \sum_{k=-K_0}^{k=K_0} \gamma_k e^{jn\omega_0 \nu_k}, \quad K_0 = (K^2 - K) / 2 \quad (11)$$

where  $\gamma_k \propto \Gamma_{kl}$  and  $\nu_k \propto t_{kl}$ . When  $K = 2$ ,  $\gamma_1 = \gamma_{-1}^* = \Gamma_0 \Gamma_1$  and  $\nu_1 = -\nu_{-1} = t_{01}$ . In practice,  $\{\Gamma_k\}_{k=0}^{K-1}$  are real-valued, but, even if this is not the case, we always have conjugate symmetry, *i.e.*,  $\gamma_k = \gamma_{-k}^*$  and  $\nu_k = -\nu_{-k}$ . We recover  $\{\gamma_k\}_{k=0}^{K-1}$  in two steps:

- ① We first estimate  $\nu_k$  given  $y_n = |\hat{m}(n\omega_0)|^2$ ,  $n = 0, \dots, N - 1$ .

Using the idea of Prony [12], we define the polynomial,

$$H(z) = \prod_{|m| \leqslant K_0} (1 - e^{-j\omega_0 \nu_m} z^{-1}) = \sum_{m=0}^{2K_0+1} h_m z^{-m} \quad (12)$$

which is yet to be computed. However, note that,  $(h * y)_n = 0$ ,

$$\begin{aligned} (h * y)_n &= \sum_{m=0}^{2K_0+1} h_m y_{n-m} \stackrel{(11)}{=} \sum_{m=0}^{2K_0+1} h_m \sum_{k=-K_0}^{k=+K_0} \gamma_k e^{jn(n-m)\omega_0\nu_k} \\ &= \sum_{k=-K_0}^{k=+K_0} \gamma_k \left( \sum_{m=0}^{2K_0+1} h_m e^{-jm\omega_0\nu_k} \right) e^{jn\omega_0\nu_k} = 0. \quad (13) \end{aligned}$$

$\Leftrightarrow H(e^{-j\omega_0\nu_k}) = 0$

Hence, the key idea is to find a filter sequence  $h$  which, when filtered with  $y_n$ , produces zeros. This problem is at the heart of *spectrum estimation* [12]. In this work, we use the composite mapping property algorithm (CMPA) devised by Cadzow [13].

- ② With  $\nu_k$  known, estimating  $\gamma_k$  in (11) is a linear problem.

We solve the linear least-squares problem to estimate  $\tilde{\gamma}_k$ ,

$$\tilde{\gamma}_k = \min_{\gamma_k} \sum_{n=0}^{N-1} \left| y_n - \sum_{k=-K_0}^{k=+K_0} \gamma_k e^{jn\omega_0\nu_k} \right|^2. \quad (14)$$

Now, since  $y_n = y_{-n}$ , we need  $N \geq K_0 + 1$  measurements to solve for  $h$  (cf. (13)) [14]. We show a step-by-step decomposition of  $y_n$  in terms of  $\{\gamma_k, \nu_k\}_{|k| \leq K_0}$  (11) in Fig. 3. Next, we discuss how to estimate  $\{\Gamma_k\}$  from  $\{\gamma_k\}$  in (14).

### 3.1. Estimating $\{\Gamma_k\}$ from $\{\gamma_k\}$ when $K = 2$

The case has two interesting interpretations. The first one is simply the setting being described in Fig. 2. An alternate interpretation arises from approximation of scattering where one sets  $K = 2$ , assuming negligible contribution of higher-order terms  $K \geq 3$ , i.e.,

$$|\hat{m}(\omega)|^2 \stackrel{(10)}{=} |\Gamma_0 e^{-j\omega t_0} + \Gamma_1 e^{-j\omega t_1}|^2 + \varepsilon_\omega(\Gamma_k, t_k)_{k \geq 3},$$

where  $\varepsilon_\omega(\Gamma_k, t_k)_{k \geq 3} = \left| \sum_{k=0}^{K-1} \Gamma_k e^{-j\omega t_k} \right|^2 \approx 0$ . In either case,

$$|\hat{m}(\omega)|^2 = \underbrace{|\Gamma_0|^2 + |\Gamma_1|^2}_{a_0} + \underbrace{2|\Gamma_0||\Gamma_1| \cos(\omega t_{01} + \angle \Gamma_{01})}_{a_{12}}. \quad (15)$$

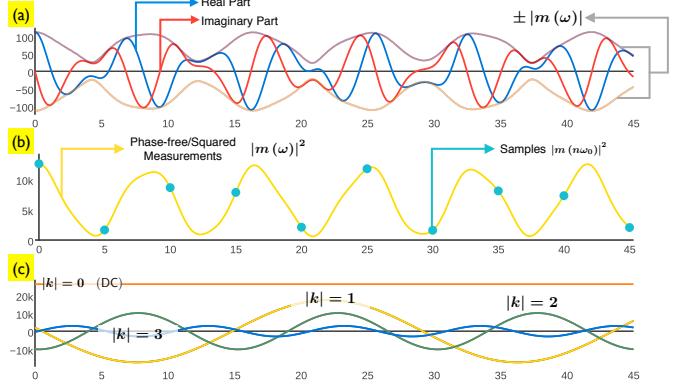
The first step towards retrieving  $|\Gamma_{0,1}|$  involves the estimation of  $a_0, a_{12}$  based on the available samples  $y_n$  (11). Given the equivalence between (15) and (11), the set of parameters  $a_0, a_{12}, t_{01}$ , and  $\angle \Gamma_{01}$  are retrieved using  $K_0 = 1$ .

The parameters  $a_0$  and  $a_{12}$  allow to estimate the values of the transmitted and reflected magnitudes  $|\Gamma_0|, |\Gamma_1|$  based on algebraic relations. Specifically, the non-negativity of  $a_0, a_{12}$  implies that  $||\Gamma_0| \pm |\Gamma_1|| = \sqrt{a_0 \pm a_{12}}$ . Thus, based on  $\tilde{a}_0, \tilde{a}_{12}$ ,

$$\left\{ |\tilde{\Gamma}_0|, |\tilde{\Gamma}_1| \right\} = \frac{1}{2} \left( \sqrt{\tilde{a}_0 + \tilde{a}_{12}} \pm \sqrt{\tilde{a}_0 - \tilde{a}_{12}} \right). \quad (16)$$

The definition of  $a_0$  and  $a_{12}$  implies the non-negativity of the above square-root argument, except if estimation errors occur. In such cases, our algorithm replaces  $\tilde{a}_{12}$  by  $\tilde{a}_0$ , thus yielding the double solution  $\sqrt{\tilde{a}_0/2}$  for both magnitude values.

Following the above estimation operations, the retrieved values must be correctly assigned to the corresponding magnitudes. The phase-less setting precludes direct identification because the absolute distance parameters cannot be retrieved as such. Now, according to the inverse-square law, one can still assume that the magnitudes of  $|\Gamma_1|$ —given the relative proximity of this map—are *in average* larger than those of  $|\Gamma_0|$ . However, this assumption does not hold for every pixel



**Fig. 3:** Phase Retrieval in Setting  $K = 3$ . This measurement corresponds to pixel  $(20, 20)$  of the synthetic experiment in Section 4. (a) Complex-valued  $\hat{m}(\omega)$  with the real and imaginary parts together with the envelope  $\pm |\hat{m}(\omega)|$ . (b) Phase-less  $|\hat{m}(\omega)|^2$  with  $y_n = |\hat{m}(n\omega_0)|^2, \omega_0 = 5$ . (c) Decomposition of  $y_n$  into terms of  $\{\gamma_k, \nu_k\}_{|k| \leq K_0}$  as in (11).

in general. This implies that the assignments may suffer from some ambiguities that cannot be solved pointwise. We thus propose to address this issue by leveraging inter-pixel scene dependencies. In particular, we exploit the fact that maps stemming from real-world scenes display a certain degree of spatial regularity. Qualitatively, the disambiguation approach that we follow consists in (a) first detecting value-crossing locations  $\mathbf{k}_i$  where both map magnitudes are (nearly) equal, and (b) then determining suitable assignments by decomposing the field of view into associated adjacent areas satisfying either  $|\Gamma_1[\mathbf{k}]| \geq |\Gamma_0[\mathbf{k}]|$  or  $|\Gamma_1[\mathbf{k}]| < |\Gamma_0[\mathbf{k}]|$ .

### 3.2. Higher-Order Settings

The magnitudes  $|\Gamma_i|$  can also be retrieved when  $K \geq 3$ . As in (16), a set of real numbers  $a_0$  and  $a_{ij}$  can be identified in association with the constant and oscillatory terms. Disambiguation even becomes more straightforward in such settings because the presence of intermediate sheets allows to assign all magnitude values of interest pointwise based on the relative-distance  $t_{kl}$ . The latter quantities are also obtained (as byproducts) from the LSE method.

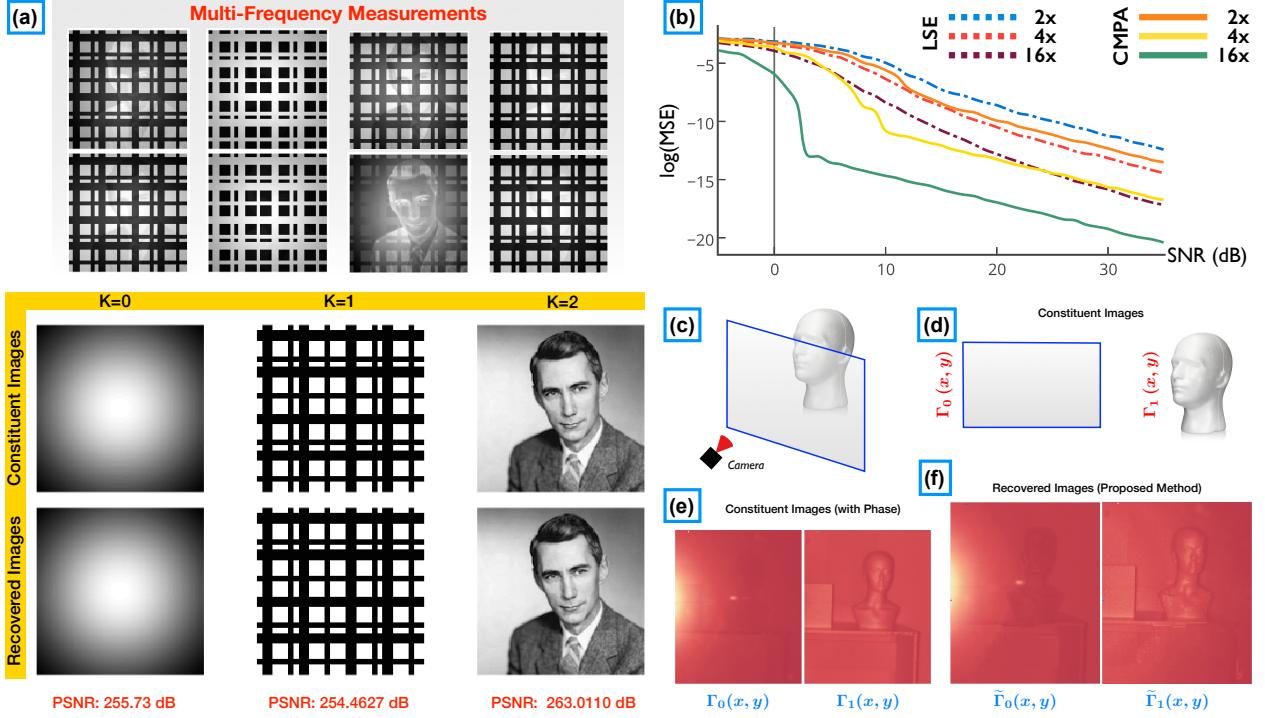
Let us consider the 3-bounce setting where the constant term is  $a_0 = |\Gamma_0|^2 + |\Gamma_1|^2 + |\Gamma_2|^2$ , the oscillatory terms being  $a_{12} = 2|\Gamma_0||\Gamma_1|$ ,  $a_{23} = 2|\Gamma_1||\Gamma_2|$ , and  $a_{13} = 2|\Gamma_0||\Gamma_2|$ . Based on the estimates of the  $a_0$  and  $a_{ij}$  provided by the LSE method, the set of values  $v_i$  corresponding to the unknown magnitudes  $|\Gamma_{0,1,2}|$  is retrieved as  $\{\tilde{a}_{12}\tilde{a}_{13}\mu, \tilde{a}_{12}\tilde{a}_{23}\mu, \tilde{a}_{13}\tilde{a}_{23}\mu\}$ , where

$$\mu = \tilde{a}_0^{1/2} [(\tilde{a}_{12}\tilde{a}_{13})^2 + (\tilde{a}_{23}\tilde{a}_{13})^2 + (\tilde{a}_{23}\tilde{a}_{12})^2]^{-1/2}.$$

In this setting, 7<sup>th</sup>-order estimation is performed with the LSE method, the order number  $(2K_0 + 1)$  being here related to the constant and to the 6 complex exponentials associated with the  $K = 3$  cosine terms (11). An estimation example is illustrated in Figure 3.

## 4. EXPERIMENTAL RESULTS

**Synthetic Data:** In this first experiment, we consider scene with  $K = 3$  with constituent images shown in Fig. 4(a). Each of the images  $\{\Gamma_k\}_{k=0}^{K-1}$  are of size  $160 \times 160$  and  $\{t_k\}_{k=0}^{K-1} = [0.31 \ 0.55 \ 0.93]\pi$ . We use  $N = 4 = (K^2 - K)/2 + 1$ , multi-frequency, magnitude measurements  $|\hat{m}(n\omega_0)|, n = 0, \dots, 3, f_0 = 1$  MHz. The corresponding results are shown in Fig. 4 (a). In this synthetic scenario, the



**Fig. 4:** (a) Experimental results for simulated data with  $K = 3$ . The images on the top are spatial maps of the available magnitude multi-frequency measurements  $|\hat{m}(2n\omega_0 + 1)|^2$ ,  $n = 1, \dots, 8$ . For  $K = 3$  we show the ground truth which consists of distinct layers constituting the scene as well as the reconstruction from the phase-less data with  $N = 1 + (K^2 - K)/2 = 4$  measurements. We list the PNSR value indicating reconstruction upto machine precision. (b) Performance analysis of the Cadzow's CMPA algorithm [13] for  $K = 2$  for the noisy case. We show the effect of over-sampling, that is  $N = \kappa(K + 1)$  samples, on mean squared error (MSE) as a function of signal-to-noise ratio (SNR). As can be seen, when  $\kappa = 2$  or we have  $2N$  samples of  $y_n$ , the CMPA algorithm has a stabilizing effect on MSE. For the case of  $16x$  oversampling, the MSE decreases linearly with increase in SNR. This effect happens as early as SNR = 3 (dB) for the CMPA algorithm. We also compare the LSE method with the CMPA method for different over-sampling factors. (c) Scene setup for experiment for  $K = 2$  with Microsoft Kinect. (d) Schematic representation of the constituent images. (e) We reconstruct the assuming that we measure  $\hat{m}(n f_0)$ ,  $n = 50, \dots, 100$  with  $f_0 = 1$  MHz. (f) Reconstruction via phase retrieval using  $|\hat{m}(n f_0)|^2$ ,  $n = 50, \dots, 100$  with  $f_0 = 1$  MHz.

reconstructions that are obtained are virtually identical to the oracle, *i.e.*, the SNR reaches machine precision. This first set of results thus already validates our approach in the synthetic, noiseless case.

The effect of noise is studied in Fig. 4(b) where we plot the observed *mean squared error* as a function of signal-to-noise-ratio (SNR) with consideration to the over-sampling factor using Cadzow's method [13]. With SNR = 20 and  $K = 2$ , 4 times over-sampling achieves reasonable bound on the MSE.

**ToF Sensor Based Data:** We use the Microsoft Kinect One ToF camera. This camera is equipped with a modified firmware facilitating customized frequencies at  $f_0 = 1$  MHz steps in range 50 to 100 MHz. The setup is described in 4 (c) and a schematic of the ground truth is shown in 4 (d). The back scene consists of a mannequin head at approximately 1.5m from the sensor. In addition, a semi-transparent sheet at a distance of 15 cm covers the sensor's field-of-view. This transparency sheet reflects a fraction of the emitted infrared light directly back to the sensor, thus  $K = 2$ . The available measurements consist of  $145 \times 119$  multi-frequency profiles captured at  $N = 51$  distinct frequencies.

Our algorithm is used for the phase-less setting where we measure  $y_n = |\hat{m}(n\omega_0)|^2$  (11). We also retain complex-valued measurements  $\hat{m}(n\omega_0)$  computed in (6). The ground truth is hard to obtain for our experiment. Consequently, we use the method described in [7, 8] to estimate  $\{\tilde{\Gamma}_0^{\text{GT}}, \tilde{\Gamma}_1^{\text{GT}}\}$  using (6) and (7) which serves as a proxy of our oracle estimate (cf. Fig. 4 (e)). Cadzow's method is employed as a stand-alone tool for this purpose. We then estimate  $\{|\tilde{\Gamma}_0|, |\tilde{\Gamma}_1|\}$  from the

knowledge of  $y_n = |\hat{m}(n\omega_0)|^2$  (11), using our proposed approach, for example, (16). As in the first experiment, the goal of these experiments is to reconstruct and identify the distinct magnitude maps  $|\Gamma_k|$  associated with  $k = 0$  and  $k = 1$ . We show the estimates obtained by our approach in Fig. 4 (f). Our results show the effectiveness of our model as well as the applicability of our algorithm. We use the SSIM index [15] as our performance metric to compare image reconstruction with and without phase measurements in Fig. 4 (e) and Fig. 4 (f), respectively. The SSIM measures for the first and the second sheets correspond to 0.750 and 0.408. Even though there is model mismatch in the data due to non-idealities of the system as well as noise, Cadzow's method [13] works reasonably well and warrants future work.

## 5. CONCLUSIONS

We have demonstrated the feasibility of image demixing from time-resolved intensity measurements. Our algorithm involved line-spectrum estimation combined with a disambiguation approach. In further work, the design of reconstruction approaches where spatial and temporal redundancy are jointly exploited could be beneficial. Since our estimates of  $\Gamma_k$  depend on the estimation of  $\gamma_k$ , which in turn depends on  $\nu_k$  (11), a rigorous analysis of the estimation method could provide new insights into studying the scattering phenomenon. Finally, we close this work by noting that, while phase-retrieval problems have been extensively studied in the literature, intensity retrieval seems to be a relatively unexplored.

## 6. REFERENCES

- [1] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Sparse ICA for blind separation of transmitted and reflected images," *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 84–91, 2005.
- [2] E. Be'Ery and A. Yeredor, "Blind separation of superimposed shifted images using parameterized joint diagonalization," *IEEE Transactions on Image Processing*, vol. 17, no. 3, pp. 340–353, 2008.
- [3] K. Gai, Z. Shi, and C. Zhang, "Blindly separating mixtures of multiple layers with spatial shifts," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008., 2008.
- [4] Q. Yan, E. E. Kuruoglu, X. Yang, Y. Xu, and K. Kayabol, "Separating reflections from a single image using spatial smoothness and structure information," in *Latent Variable Analysis and Signal Separation*. Springer, 2010, pp. 637–644.
- [5] N. Kong, Y.-W. Tai, and S. Y. Shin, "High-quality reflection separation using polarized images," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3393–3405, 2011.
- [6] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: a survey," *IEEE Sensors Journal*, vol. 11, no. 9, pp. 1917–1926, 2011.
- [7] A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar, "Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization," *Optics Letters*, vol. 39, no. 7, 2014.
- [8] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, and R. Raskar, "Resolving multipath interference in Kinect: An inverse problem approach," in *IEEE Sensors*, Nov 2014, pp. 614–617.
- [9] A. Bhandari, A. Kadambi, and R. Raskar, "Sparse linear operator identification without sparse regularization? Applications to mixed pixel problem in time-of-flight / range imaging," in *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, 2014, pp. 365–369.
- [10] A. Bhandari, C. Barsi, R. Whyte, A. Kadambi, A. J. Das, A. Dorrington, and R. Raskar, "Coded time-of-flight imaging for calibration free fluorescence lifetime estimation," in *Imaging Systems and Applications*. Optical Society of America, 2014, pp. IW2C–5.
- [11] J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, "Fourier analysis on transient imaging by multifrequency time-of-flight camera," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014, p. to appear.
- [12] P. Stoica and R. L. Moses, *Introduction to spectral analysis*. Prentice Hall, Upper Saddle River, 1997, vol. 1.
- [13] J. A. Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.
- [14] G. Plonka and M. Wischerhoff, "How many Fourier samples are needed for real function reconstruction?" *Journal of Applied Mathematics and Computing*, vol. 42, no. 1-2, pp. 117–137, 2013.
- [15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.