Compressive Acquisition and Computational Reconstruction of Ray Space

by Qiang YAO

Department of Electrical Engineering and
Computer Science

Graduate School of Engineering

Nagoya University

A thesis submitted for the degree of

Doctor of Philosophy

Sep. 2014

I would like to dedicate this thesis to my loving wife.

Acknowledgments

I would like to first express my great gratitude to my supervisor, Professor Toshiaki Fujii, for having me as his student and providing me the opportunity to pursue the research of ray space that led to this thesis. Due to his great help and continuous support, I have not only learned how to do but importantly learned how to think as well in fighting against numerous difficulties, solving countless problems and conquering various failures. I deeply cherish the time I have spent working with him, which is the most unforgettable memory in my life. His enthusiasm, love and hard working attitude as a scholar has always motivated me to pursue research as a career.

It is a great fortune for me to have Associate Processor Keita Takahashi in the lab as my great mentor, and I sincerely appreciate the discussions with Professor Takahashi. His great passion, rigorous attitude and smart intelligence in research have always been my great sample to respect and learn.

Thanks to Processor Ken-ichi Sato and Associate Processor Daisuke Deguchi for reviewing this thesis and their insightful comments. I would also like to thank Dr. Mehrdad Panahpour Tehrani for his great insights, suggestions, kind encouragement and help in Fujii Lab. In addition, I would like to show the great respect and appreciation to the Professor Masayuki Tanimoto who has been retired from Nagoya University.

I am grateful to the graduated Dr. Yanlei Gu who has provided his warm-hearted helps in my daily life and study as I came to Japan. Thanks especially to the secretary Makiko Abe and my tutor Keita Maeno who provided me great helps to get through many difficulties. Many thanks to all current and former researchers and students for making my time at Fujii Lab such a pleasure.

Finally, I wish to thank my parents and my wife, whose love makes it all worthwhile.

Abstract

Ray space (light field in other literatures) has been regarded as one of the most exciting medias in image representation and computer vision. Ray space records the whole information of lights that pass through one real space, and the applications of ray space have become more popular and more attractive, which include free viewpoint image (FVI) generation for watching different perspective images, all-in-focus images generation for image refocusing, and compressive display. However, the data amount of a ray space is quite huge, and this problem is enhanced by the requirement of higher image resolution and denser sampling rate of ray space. Therefore, in this thesis, we investigate the compressive acquisition and computational reconstruction of ray space by adopting the conception of compressed sensing, which states that signal can be exactly reconstructed from incomplete measurements obtained by incoherent projection if the signal is sparse by itself or sparse in other domain. The basic target of our work is to simulate the compressive acquisition process of ray space, and evaluate the relation between the reconstruction quality and the number of sensed measurements. The simulation results can provide a good reference in the design of optical acquisition system. Based on this target, we also attempt to attain better reconstruction quality at a fixed sensing ratio by proposing new methods, which are regarded as the main contributions of this thesis.

The first contribution relates to the combination of sparse coding and reconstruction of the ray space. After establishing the acquisition model and obtaining the compressively sampled measurements of a ray space, in the reconstruction procedure, we propose to reconstruct the ray space from compressively sampled measurements by exploring sparsity of the ray space, where epipolar plane image (EPI) is adopted as the processing unit. In order to obtain sparser representation of EPI, two types of dictionaries are proposed. Normally, Gabor functions can be used to encode the motion compensated frames in video processing, and similarly EPI actually reflected the disparity information between each free viewpoint image (FVI). Therefore, the dictionary is carefully

designed using Gabor function by tuning different parameters, so that the generated atoms in 2D Gabor dictionary can match the features of EPI. As a result, sparser representation of EPI can be obtained. In the experiments, we also take two other orthogonal bases to make comparison of the reconstruction quality. The results show that better reconstruction quality is achieved by 2D Gabor dictionary. The success of adopting overcomplete dictionary (whose number of atoms exceeds the dimension of the dictionary space) in representation and reconstruction of ray space promotes us to continuing exploring better dictionary by more sophisticated method.

Therefore, another dictionary is proposed by adopting dictionary learning method, and the dictionary can be adaptively amended from a set of training data, thus the atoms obtained from dictionary learning can be shaped to grasp features of EPI. In the experiments, we compare the 2D Gabor dictionary, the learned dictionary, and another orthogonal basis. The best reconstruction quality of ray space is achieved by the learned dictionary, followed by the 2D Gabor dictionary, and the orthogonal basis gets the worst result.

However, the two dictionaries mentioned above sacrifice reconstruction speed to gain high reconstruction quality, and the slow reconstruction speed becomes an obstacle for real time applications. Therefore, in order to get a faster reconstruction speed while preserving the good reconstruction quality, we propose a statistically weighted model and integrate the model into optimization for sparse solution, which is regarded as the second contribution.

The second contribution attempts to handle the trade-off between reconstruction quality and speed. We find that the amplitude structure of coefficients of ray space in DCT domain also provides another piece of priori information in the reconstruction, and we propose designing a weighted matrix to reflect the structure and to integrate the structure in the reconstruction process, which is operated by l_1 norm optimization. In addition, we provide a solution for the new optimization problem so that the previous optimization solver can be reused. The experimental results show that the weighted model-based method achieves better reconstruction quality and faster reconstruction speed than the conventional method, plain l_1 norm optimization. Furthermore, the weighed model-based method and dictionary learning-based method achieve similar performance (much better than the conventional method) in the aspect

of reconstruction quality. However, the weighed model-based method has great advantage in reconstruction speed. Therefore, the second contribution obtains faster reconstruction speed while preserving high reconstruction quality.

Based on the contributions mentioned above, the compressively sampled ray space can be well reconstructed by computational methods. Even at quite low sensing ratio, such as 10% of the total pixels in one ray space, the main part of ray space can still be recovered. As the sensing ratio increases to around 50%, the reconstruction quality of ray space is quite promising. In addition, considering the reconstruction time, the second contribution can greatly accelerate the reconstruction speed, because the addition of weight promotes the convergence in the iteration for estimation of sparse solution. Therefore, the huge data problem of ray space in acquisition stage is resolved, and the ray space can also re constructed well by the proposed computational reconstruction methods.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we start with the motivation and goal of this work. The high demand from academic and practical areas of ray space acquisition have convinced ourselves that all the work we are doing is necessary and imperative. Following the explicit target, our contributions in this thesis are clarified and they will compose the main body of this thesis. In the final part of this chapter, the whole structure of the thesis is given out, and the relationship between each chapter is also visualized.

1.1 Motivations and Goals

"Let there be light," God said. Then, there was light. In ancient time, light was solely from the sun. Later, it could also be from fire. As Thomas Edison claimed that the first commercially practical incandescent bulb had been invented [1][1], light could be created artificially and applied on a large scale commercially. However, can we capture light?

Light capturing has always been regarded as a dream of human beings. As early as in 1826, the first camera was invented by Joseph Nicéphore Niépce to record the light information permanently [2]. From that moment, visual information acquisition became realistic. Particularly, since the middle of the last century, the development of computer technology and advance in LSI technologies take us into the digitalized era. Following

[1]Edison did not invent the first electric light bulb, but instead invented the first commercially practical incandescent light.

the trend of digital explosion, digital camera has attained a significant advancement; from film to digital sensor, from monochrome to color, from a large machine for professional photographers to a compact toy for a large number of consumers. Camera is becoming smarter and more powerful, however, the basic principle of photography remains almost unchanged. What we have captured is still only a projection of light on sensor, named as image rather than whole light (ray space). [2]

Actually, independently from the development of image acquisition by traditional camera, the earliest work trying to capture ray space can be dated back to the beginning of the last century. Ives invented parallax barriers [3] and Lippmann proposed lenslet arrays [4] based on the technique of integral photography, which were initially dedicated to glass-free stereoscopic perception. Since then, however, this conception and the dream of ray space acquisition have been sealed over half a century.

Until the last decade of the 20th century, people in the field of computer vision started to reconsider ray space acquisition by utilizing computational methods. Thus, "computational photography" was born, and grew quickly. Computational photography appeared around the blurred boundary among optical light design, signal processing, and computer vision, etc., and it is regarded as a general conception to integrate computing, capturing, modulating, analyzing, processing the visual information for various applications [5]. The key of computational photography is the conception of plenoptic function [6, 7], which is a fundamental tool in the description and analysis of ray space. Therefore, ray space can be parameterized and become more achievable for us, resulting in a number of research results and academic publications. **Figure 1.1** shows the publication number when the "computational photography", "light field" or "ray space" are input as key words in Google scholar searching engine.

Along with the research on computational photograph, the appearance of numerous applications of ray space served as catalyst to promote the

[2] Ray space has the same meaning as light field in other literatures. Both of ray space and light field represent same parameterization for a real space where a bunch of lights penetrate through. Thus, in the rest part of this thesis, we prefer adopting ray space as the name of whole information of light.

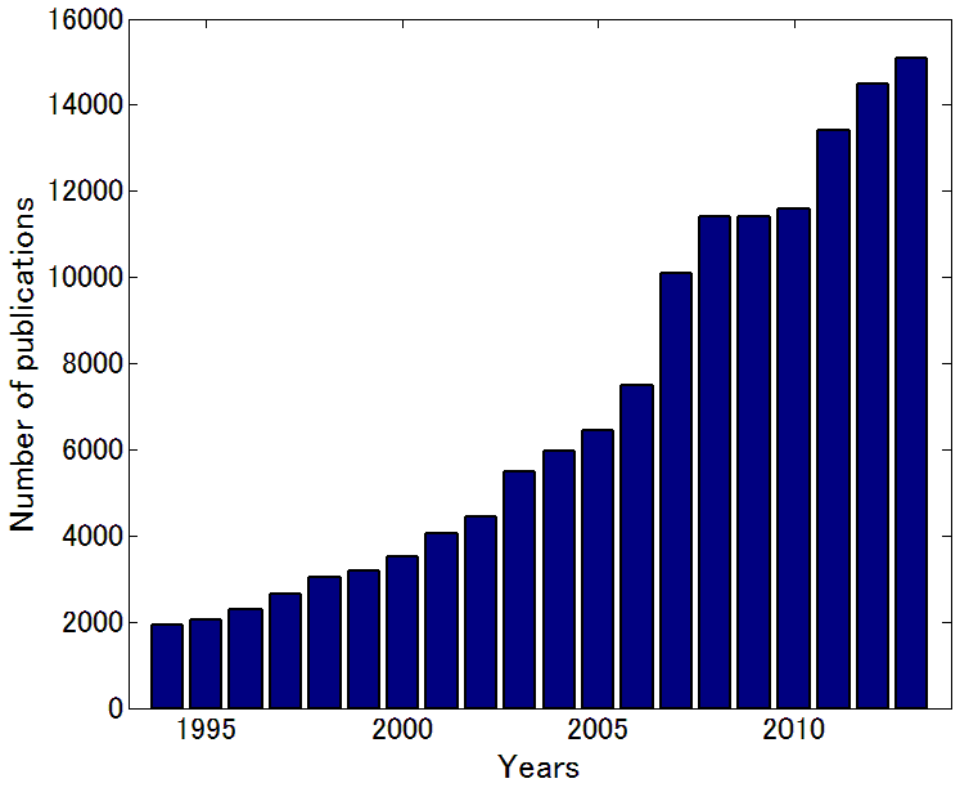


Figure 1.1: The number of publications relating to computational photography and light field

researches on ray space and its acquisition. For instance, depth information can be abstracted from ray space, which is of great significance in 3D video reproduction and processing [8–11]. In addition, free viewpoint image (FVI) [12–15] can be rendered from acquired ray space by image-based rendering [16–19] since all the geometric information is included in ray space. Besides, all-in-focus images generation [20–23] enables consumers to take photograph firstly and adjust the focus point arbitrarily in the post processing, which might seem to be a revolution for camera photograph.

However, due to the highly dense sampling requirement of ray space [24] and the increasing demand for high resolution of photograph, another

challenge is brought in ray space acquisition, which is huge data problem. Not only does the problem exist in data storage and transmission, but the problem happens at the capturing stage. It is not affordable to have such huge amount of data in acquisition, even worse, it is impossible to capture all the data in a certain condition.

Therefore, compressive acquisition is not only a practical method in capturing ray space, but also sometimes is a must in sensing ray space. In the other side of this whole chain, reconstruction is also of great significance. We are motivated to conduct compressive acquisition, and in addition we are required to recover the whole ray space from the compressively acquired measurements by computational method. This work is also motivated and inspired by the appearance of compressed sensing. [25–28].

Actually, compressed sensing is rather a principle than a methodology, and it has gained a great popularity in signal processing area due to its surprising and interesting theoretical nature and wide area of applications. It is stated in compressed sensing theory that a signal can be reconstructed quite accurately from a relatively smaller number of measurements than the total number required by traditional methods, such as Nyquist Sampling principle, as long as that the signal is sparse (compressible) in certain transform basis or dictionaries. [29,30] Besides, compressed sensing also requires the incoherence between sensing process and sparsely representing process. In fact, compressed sensing is still quite young and it was born just in 2006. However, there was an explosive beginning for compressed sensing with steady growth. **Figure 1.2** shows the publication number when the "compressed sensing" or "compressive sensing" is input as a key word in Google scholar searching engine. In addition, the scope of application of compressed sensing is still expanding. Therefore, it is interesting, necessary, and natural to have a handshake between ray space acquisition in computational photography and compressed sensing in signal processing.

In fact, the theory of compressed sensing only deals with the reconstruction possibility and necessary conditions, leaving an open and wide space for method and algorithm development in the stage of acquisition and reconstruction.

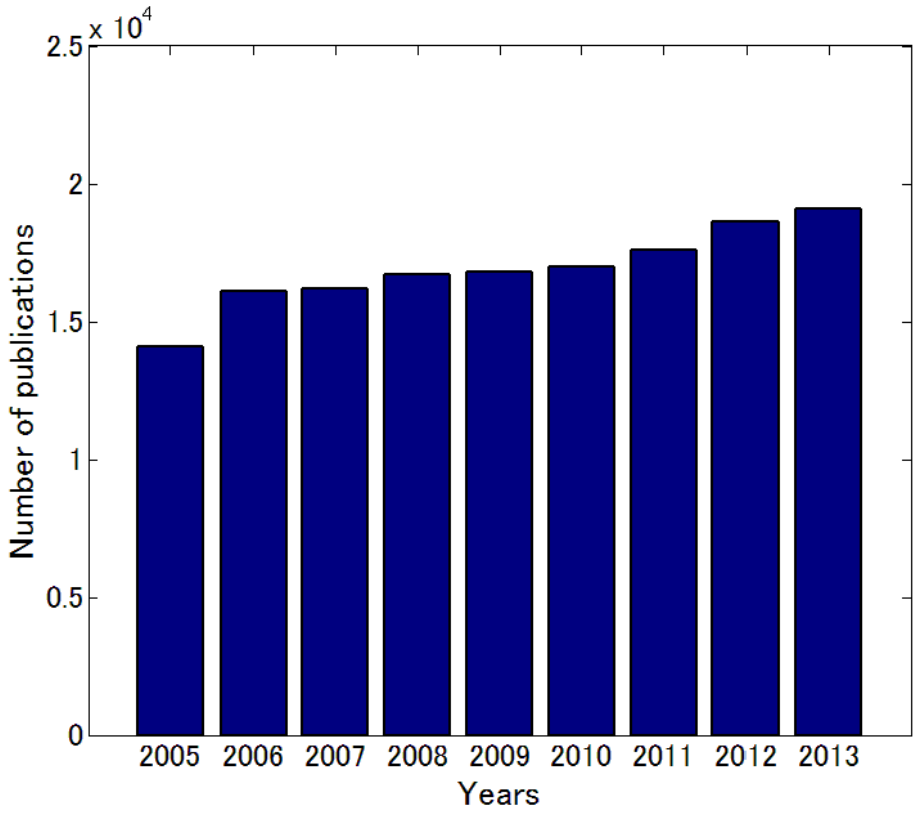


Figure 1.2: The number of publications relating to compressed sensing or compressive sensing

The basic target of this work is to analyze how compressively the ray space can be acquired and to provide the simulation results on how well the compressively acquired ray space can be reconstructed. Therefore, in this thesis, the main work is focused on the reconstruction from highly incomplete samples that are obtained by compressive sensing method. First of all, this work provides a simulation procedure, which means the ray space data have already obtained. Thus, we started from reconstruction part, and tried to improve the reconstruction quality given certain compressive ratio in acquisition. The work in this thesis is different from data compression although both of them seem to share the similar target. The largest difference is that total data acquisition is a must for data com-

pression while we try to reduce the data amount from the acquisition, and this point is also the great superiority of our work. In the second place, the simulation has to be dedicated to the practical implementation. Thus, both of the reconstruction quality and speed have to be improved simultaneously. Based on those specific targets, our main contributions are developed.

1.2 Contributions

In this thesis, the main contribution is that compressive acquisition of ray space is proposed and the quantitative relationship between the compressive sensing ratio and reconstruction quality is explored as well. In addition, two computational reconstruction methods are proposed to enhance the reconstruction quality of ray space.

Generally, the contribution of this thesis relates to several fields including image processing, sparse coding, optimization problem, computational photography and application of compressed sensing. To be specific, these contributions can be attributed to the following two parts.

First of all, sparse representation is one of the necessary conditions for attaining high quality reconstruction from incompletely sampled ray space. Since a ray space is neither sparse in spatial domain nor in angular domain, but quite redundant in both domains, we made effort to find sparse representation of ray space in other domain. Thus, several orthogonal bases were chosen for sparse representation as the benchmarks of the research. Moreover, we realized that the reconstruction quality could be greatly enhanced if ray space could be more sparsely represented. Therefore, we proposed to design one dictionary to grasp the feature of ray space, and demonstrated sparser representation of ray space by the designed dictionary. Furthermore, we proposed to integrate dictionary learning into ray space reconstruction, because a well-learned dictionary could provide much sparser representation of ray space than a carefully-designed dictionary. The illustrated results showed that learned dictionary enjoyed the best performance in terms of reconstruction quality. Therefore, this contribution can be viewed as novel and important combination of sparse

representation of dictionary and reconstruction of ray space.

Secondly, although it has been shown that the extension from an orthogonal basis, such as a discrete cosine transform (DCT) domain, to overcomplete dictionary can bring sparser representation, this point is only based on the priori information of signal sparsity. Actually, however, there are other priors of ray space representation, and if the additional prior can be exploited, better reconstruction quality can also be expected. Therefore, in the second contribution, we found that the distribution of non-zero elements showed the structure in the orthogonal basis rather than the learned dictionary, and this kind of structure was utilized in optimization process by the proposed model to estimate more accurate sparse solution, so that better reconstruction quality was achieved. Besides, the computation time was greatly reduced since this method was based on the orthogonal basis. Thus, this contribution can be viewed as combination of image processing and optimization problem.

All in all, the two contributions mentioned above illustrated that the ray space could be compressively acquired and computationally reconstructed. In addition, we also evaluated the quantitative relation between reconstruction quality and the necessary numbers of acquisition. Therefore, these contributions could provide good support and reference for the future optical design, which is regarded as one of the next targets of our research.

1.3 Outline of Thesis

The rest of this thesis is organized as follows.

Chapter 2 Background and Related Works. In this chapter, we firstly present the background of this work, which is regarded as a merge among several fields. Based on the background, we also present an overview of the related works.

Chapter 3 Compressive Acquisition and Reconstruction of Ray Space by Sparse Representation of Dictionary. In this chapter, we adopt sparse coding to obtain sparser representation of ray space, and propose two types of dictionaries developed with parameter tuning

and dictionary learning, respectively. The experimental results show that the reconstruction quality is greatly improved by adopting overcomplete dictionaries.

Chapter 4 Reconstruction of Ray Space by Statistically Weighted Model. In this chapter, we consider the amplitude structure of coefficients of ray space in DCT transform domain, and design one weighted model statistically from a large training data set. This model is integrated in the estimation of sparse solution. Thus, the reconstruction quality is greatly improved and reconstruction time is reduced.

Chapter 5 Conclusion and Future Works. In this chapter, we conclude the whole work in this thesis, and discuss the limitations of the current work. Based on these limitations, several future perspectives are presented.

The organization diagram of the thesis is illustrated in **Figure 1.3**.

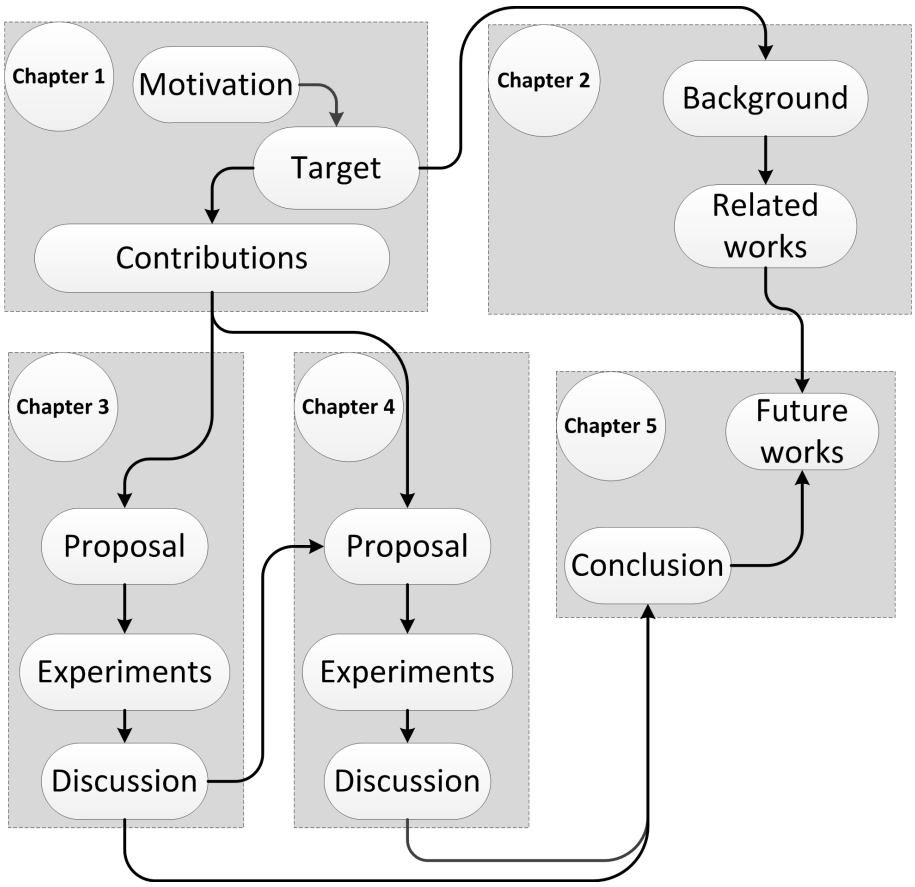


Figure 1.3: The relationship between each chapter in the thesis

Chapter 2

Background and Related Works

In this section, the background in where this research is rooted, and the basic knowledge are illustrated firstly. The background of this work is basically composed of three parts, which are description of ray space, application of ray space, and theoretical support for compressive acquisition of ray space.

The related works in acquisition and reconstruction of ray space are divided into three parts, consisting of l_p norm optimization ($p \leq 1$), acquisition systems and mathematical model for acquisition. The proposals and contributions in the following chapters are mostly based on these related works. Next, we describe the background and related works in detail.

2.1 Background

The description and parameterization of ray space enable readers to have a vivid impression of ray space. The feasible applications of ray space illustrate the significance of our research. In addition, the theoretical support of the research, compressed sensing, is also briefly introduced as the background, so that it will be much easier for readers to understand the rest contents and our contributions.

2.1.1 Plenoptic Function and Ray Space Parameterization

The study of light acquisition started from several centuries ago, however, there was few specific descriptions of light or ray space until the appearance of plenoptic function, [1] which was firstly proposed by Adelson and Bergen in 1991. [6]. In general, a ray space is described as the radiance density function on the final image sensor under the assumption that the scene is Lambertian or intensity invariant from any directions. Parameterized by plenoptic function, the whole information of lights in the ray space is recorded. To be specific, there are basically seven parameters in plenoptic function for the description of the radiant energy, written as

$$\mathcal{P} = P(x, y, z, \theta, \varphi, \lambda, t). \tag{2.1}$$

The intensity of $\mathcal{P}$ denotes the value of energy that is perceived at any 3D viewing point (x, y, z), from the direction (θ, φ) for the light with wavelength λ at the time t. If we only discuss the visible lights, the parameter λ is fixed. Besides, if we only take static photograph into consideration, the parameter t becomes a constant. Thus, the rest five parameters, $P(x, y, z, \theta, \varphi)$, where $-\pi \leq \theta \leq \pi$, and $-\pi/2 \leq \varphi \leq \pi/2$, can fully describe the information of light of a real scene, as shown in **Figure 2.1**.

Moreover, if the reference plane (or we name it camera plane) is fixed during the image acquisition, the parameter z can be further omitted, as shown in **Figure 2.2(a)**. Thus, what we can capture is the spatial information, (x, y) and the angular information (θ, φ). This representation can be equivalent to a two-plane data parameterization, as shown in **Figure 2.2(b)**. The two planes include one camera plane (sensor plane) parameterized by (u, v) and a reference plane parameterized by (s, t) with the distance f. If we set

$$u = x, \tag{2.2}$$

[1] The "*plenoptic*" was named by Adelson and Bergen by using the latin root *plenus* with the meaning of complete or full and *optic* with the meaning of vision.

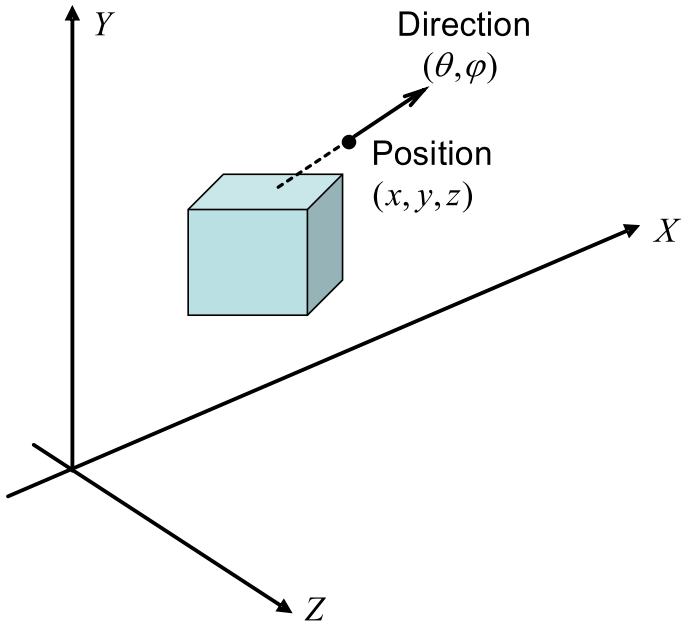


Figure 2.1: 5D parameterization for a ray space

$$v = y, \tag{2.3}$$

$$s = x + ftan\theta, \tag{2.4}$$

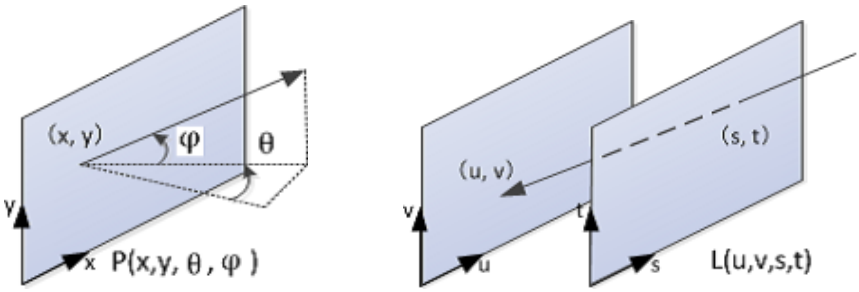
$$t = y + ftan\varphi/cos\theta, \tag{2.5}$$

the parametrization of 4D ray space is transfered to the parameterization of two-plane data, which completely describes the spatio-angular light variation in a space without considering occlusion. If we define $(u, v) \in \Gamma$ and $(s, t) \in \Pi$, there is

$$L : \Gamma \times \Pi \rightarrow R, \quad (u, v, s, t) \mapsto L(u, v, s, t) \tag{2.6}$$

There are several names for this kind of parameterization, such as light field, [31], Lumigraph, [32] or ray space, [33] and we prefer adopting ray space through the whole thesis. The two-plane parameterization $L(u, v, s, t)$ will be employed in the rest contents of this thesis as the description of a ray space.

In the next place, we introduce another image format, named as epipolar plane image (EPI) [34], which is of great significance in analysis of ray



(a)4D ray space (b) two-plane ray space

Figure 2.2: The equivalence of 4D ray space and two-plane ray space

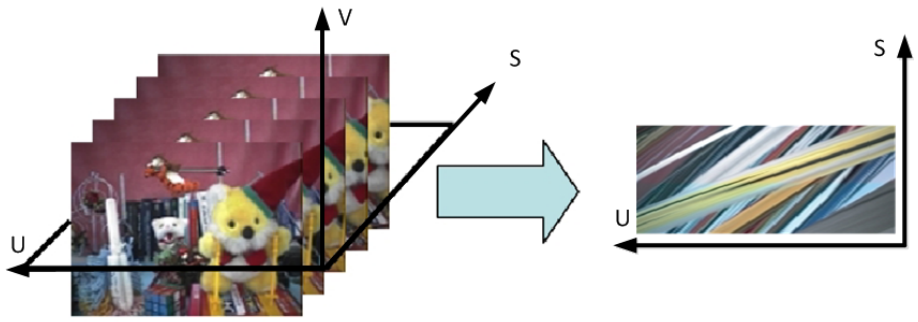


Figure 2.3: The illustration of EPI generation from ray space

space and structure description of scene, and EPI is also the basic processing unit for reconstruction in the next chapter.

Since a ray space can be parameterized as $L(u, v, s, t)$, we fix a horizontal line of constant t' in focal plane and a constant v' in camera plane, thus the ray space is restricted to a sliced plane $\Sigma_{t',v'}$, or represented as (u, s). This sliced image is named as epipolar plane image (EPI), represented as

$$(u, s) \mapsto \Sigma_{t',v'}(u, s) := L(u, v', s, t'), \tag{2.7}$$

since it includes the geometric information of ray space. The illustration of EPI is shown in **Figure 2.3**.

Let us consider a much simpler case as shown in **Figure 2.4** for analysis. Assume there are two points, A and B, inside of the ray space $L(u, v', s, t')$, and the distance between two planes are f. Besides, the

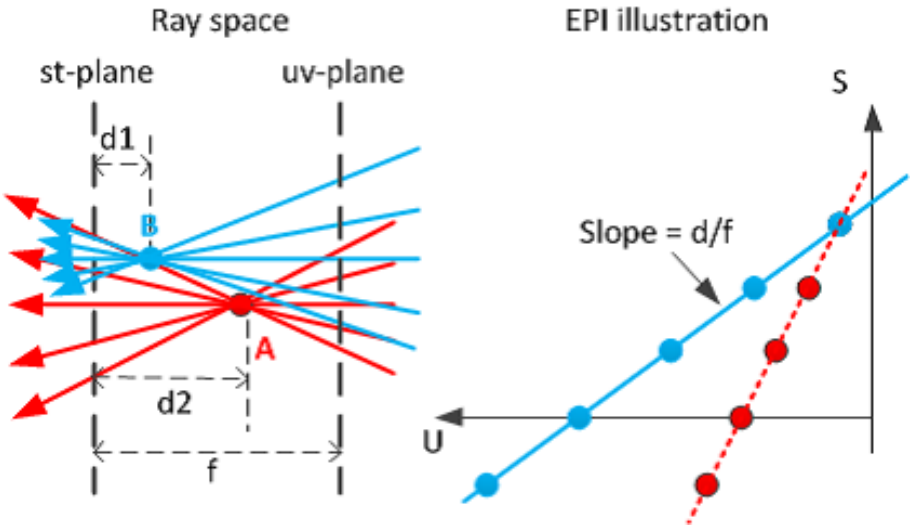


Figure 2.4: The analysis of EPI

distance from each point to st-plane is d_1 and d_2 respectively. Then, we have

$$\triangle s = \frac{d}{f} \triangle u, \tag{2.8}$$

where $\frac{d}{f}$ is the slope of each line in EPI illustration. It can be found that the point that is closer to st plane has smaller slope in EPI while the point that is farther from st plane gets larger slope in EPI. In addition, the line with smaller slope covers ahead of the line with larger slope, because objects with smaller depth occludes objects with larger depth. Therefore, EPI actually represent depth information of scene and occlusion relation can also be inferred from EPI. More significantly, EPI owns certain structure that it is composed of straight lines with different slopes. Thus, it provides us an opportunity to grasp this structure in certain domain and obtain sparser representation of EPI. The detail will be mentioned and discussed in the following chapter.

2.1.2 Computational Photography and Applications

Our work mainly focuses on acquisition and reconstruction of ray space, which relates closely to the conception of computational photography (CP) [35, 36]. Actually, computational photography is a large conception that integrates computation, digital sensor, optics, signal processing, reconstruction, and etc., together. It tries to obtain richer visual perception, captures more information that is far beyond just simple pixel level, and makes the recorded scene more flexible for varied applications. In general, computational photography differs from traditional film-like photography in several aspects. [37, 38]

(i) Generalized Optics: There is a classical optic system for traditional photography to capture the information of position and color. However, in computational photography, it enables to capture information of different wavelength, such as spectral imaging [39, 40], or gamma-ray and X-ray astronomy by coded aperture imaging. [41, 42]

(ii) Generalized Sensing: As we mentioned in previous chapter, traditional photography can only capture the 2D projection of ray space. However, computational photography attempts to get more, and obtaining a 3D or 4D ray space is the target. For instance, the Plenoptic Camera [43, 44] subdivides the integrated rays into separate measurements, so that image refocus and perspective navigation become possible.

(iii) Generalized Reconstruction: For traditional photography, demosaicking process is performed, perhaps followed by low-pass filtering, and then the captured image can be reconstructed. However, computational photography can do more, and the captured measurements can be any other data by considering camera intrinsic parameters and particular applications. The reconstruction is also complicated, and sometimes is quite time consuming if optimization is required. The general reconstruction property convinces us that it is possible and feasible to reconstruct the entire ray space from fewer sampled measurements by computational method. [45–48]

The appearance of computational photography is also partly resulted from our requirement of light information and acquisition motivated by

different applications. We do not plan go deeper in computational photography, but we would like to point out several practical applications of ray space. For one thing, it shows that our work does not seem to be meaningless, and for another thing, readers can have more vivid impression to our work in the following chapters.

One of the most popular applications of ray space is the generation of free viewpoint image (FVI), [12–15] and FVI is one of the most distinguished representatives of the next generation of multimedia and is also one of the hottest topics in the field of image processing and multimedia application. For the generation of FVI, 'cutting' a ray space is a simple and straightforward method, as illustrated in **Figure 2.5**, and it is not necessary to obtain the geometric information of objects and scenes. In the acquisition procedure, the rays which pass through the real space are densely sampled by multiple camera system or other sophisticated optical devices. In the display stage, the ray-space data are extracted perpendicular to the $x-u$ plane and displayed on the screen of television or mobile devices. This elegant characteristic favors us to synthesize arbitrary viewpoint images fast without any complicated models or rendering process, so that the real-time implementation becomes possible. For instance, as shown in Fig.2.5, section 1 and section 2 corresponding to different viewpoints are generated by simply cutting the ray space.

Next, another important application of ray space is the generation of all-in-focus image [20–23]. As the aperture of lens becomes larger, we can capture more lights to enhance the signal to noise ratio (SNR) in image sensor. However, larger aperture can also result in blurring effect out of focus areas. From the viewpoint of art, creamy and smooth bokeh is good for photography, however it will be a disaster for object recognition because several areas are no longer recognizable. For instance, in printed circuit board (PCB) detection, a small blurring area will miss an significant circuit part. In addition, all-in-focus image enables image refocusing, which is impossible for traditional film-like photography. As shown in **Figure 2.6**, the focus point of left image is in background and we can see the Japanese characters clearly, while the right image is focused at the nose of the yellow bear and hence the face of bear becomes quite sharp. Please

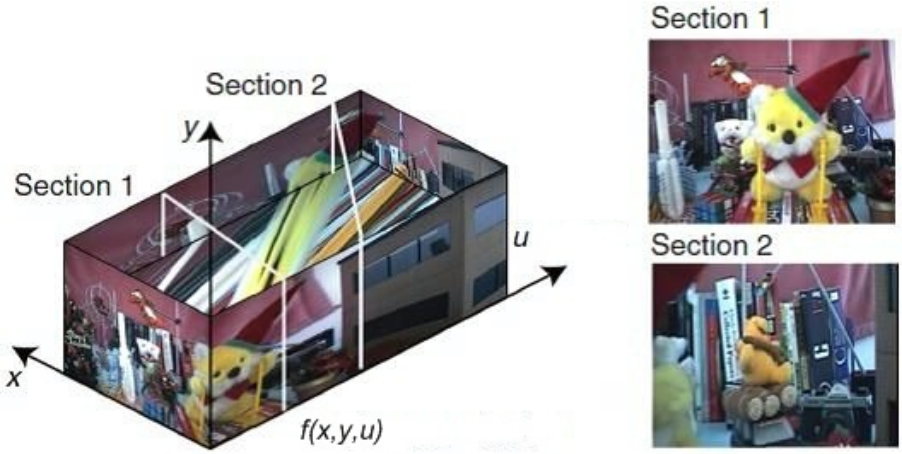


Figure 2.5: The generation of free viewpoint image (FVI) by 'cutting' ray space

note that these two images are generated from all-in-focus image, rather than taking twice shots with different focusing points. Furthermore, all-in-focus images also enable free perspective navigation, which means free viewpoint image can also be synthesized from all-in-focus images.

The last but not the least application of ray space is the compressive display [49, 50]. Different from FVI that sends the whole ray space to display, compressive display makes use of perceptual limitations of the human visual system to produce the approximation that is almost indistinguishable from the ground truth at much lower hardware complexity and cost. In addition, since the hardware for compressive display has insufficient degrees of freedom to precisely represent the whole content, it relies on an optimization procedure to determine parameters for the perceptually acceptable approximation. Recent proposals in this area include but not limited in low-rank light field displays [51], tomographic multilayer displays [52], tensor displays [53], and adaptively compressive display [54]. One illustration of compressive display is shown in **Figure 2.7**, and we can feel different viewpoints by switching different positions in front of display.

In fact, compressive display and compressive acquisition of ray space

Figure 2.6: Image refocusing by all-in-focus image

Figure 2.7: The illustration of compressive display by switching different viewing positions

are a pair of dual problem. For compressive acquisition of ray space, we start from nothing, trying to capture the whole information of a ray space. In contrast, for compressive display, we start from an entire ray space, and try to abstract certain parts and parameters that are necessary in synthesizing all information of the ray space. Then, the parameters are sent to the display for generating the entire ray space, and autostereoscopic 3D perception from any positions can be achieved.

2.1.3 Theoretical Support - Compressed Sensing

After mentioning the descriptions and applications of ray space, it is necessary to introduce the theoretical support of our research, compressed sensing, which provides the guarantee for compressive acquisition and reconstruction of ray space. Before the birth of compressed sensing, the theoretical foundation of signal acquisition is based on so-called *Shannon-Nyquist Sampling Principle*, [55] which demonstrated that signals, images, videos and other data could be exactly recovered from a group of uniformly spaced samples at the rate of twice highest frequency of the interested signals. Due to the sampling principle, digitalization has been enabled and much of signal processing has moved from the analog to the digital domain. However, sampling theory is a double-edge sword. In many significant applications, the required Nyquist rate is so high that we need to capture too many samples. More importantly, such high sampling rate may be too costly, or even physically impossible at certain circumstances to build the devices that have the capability of collecting data at the necessary rate. Furthermore, even if the data can be sampled and obtained, they are quite redundant and compression is necessary before transmission and storage. In other words, it costs a lot to sample so many points, but most of them will be thrown away in compression stage.

Compressed Sensing (CS) is an exciting, swiftly growing researching field which has obtained great attentions, achievements, and applications in signal processing, applied mathematics, statistics and computer science [25–28]. CS offers a novel framework for sampling and compressing finite dimensional data simultaneously. In addition, surprisingly, it is pointed out and proved that the high dimensional data can be recovered from incompletely sampled low dimensional measurements with overwhelming high probability as long as certain conditions are satisfied.

One necessary condition is named sparsity [56], while the other one is named the restricted isometry property (RIP) [57] or coherence. Before mentioning the two necessary conditions, we define the vector space for signals, which means any finite dimensional signal can be represented as one vector in corresponding vector space. Therefore, any signal y is written

as $y \in R^N$, and it can also be interpreted as there are N elements in signal y.

Sparsity: Normally, one natural signal can be well approximated as a linear combination of several elements from a given basis or dictionary. Mathematically, the representation is written as

$$y = \sum_{n=1}^{k} a_i \varphi_i, \tag{2.9}$$

where a_i are the coefficients and φ_i are elements in basis or dictionary, $\varphi_i \in R^N$. Besides, $k \leq N$, which means the number of elements is supposed to be no larger than the dimension of element in y. If $k < N$, we say that the signal y is sparse signal or $k-$sparse signal. However, an important point in practice is that few natural signals are truly sparse signals, rather they are compressible, which means those signal can be well approximated by sparse signal with a relatively small distortion. Mathematically, the representation is written as

$$y_c = y + \varepsilon, \tag{2.10}$$

where y is sparse signal and ε is an extremely small non-zero value. To be general, we define one matrix Φ as compressed matrix, where $\Phi = [\varphi_1, \varphi_2, ..., \varphi_M]$. Besides, $A = [a_1, a_2, ..., a_M]^T$, which corresponds to coefficient vector. In comparison to Eq. (2.9), k represents the number of elements in Φ for representing y, while M is the total number of elements in Φ. Clearly, M is supposed to be no smaller than N. If the compressed matrix Φ indicates a basis or dictionary, as $M = N$, Φ is basis while as $M > N$, Φ is an overcomplete dictionary. Therefore, the sparse representation of y_c can be formulated by matrix multiplication, written as

$$y_c = A\Phi + \varepsilon. \tag{2.11}$$

We say that signal y_c is compressible signal. Therefore, one necessary condition for compressed sensing is that the signal is supposed to be either sparse or compressible.

Restricted isometry property (RIP): Next, we define another significant matrix Ψ, named as sensing matrix. [2] Thus, the combination of sensing matrix and compressed matrix, written as $D = \Psi\Phi$, has to satisfy RIP. In [57], authors stated the following isometry condition on matrix D and established its significance in compressed sensing theory.

A matrix D satisfies the restricted isometry property (RIP) of order k if there is a $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k) \parallel \theta \parallel_2^2 \ \leq \ \parallel D\theta \parallel_2^2 \ \leq \ (1 + \delta_k) \parallel \theta \parallel_2^2 \tag{2.12}$$

holds for all $\theta \in \Sigma_k$, where $\parallel \bullet \parallel$ represents l_2 norm.

In other words, if a matrix D satisfies the RIP of order $2k$, formula Eq. (2.12) can be explained as saying that D approximately holds the distance between any pair of k-sparse vectors. Therefore, we say that a signal y can be recovered from incompletely sampled measurements with high probability as long as the signal enjoys k-sparse representation in compressed matrix Φ and the sensing matrix Ψ is well designed so that $\Psi\Phi$ can satisfy the RIP of order $2k$.

Next we consider how many measurements are sufficient to get good recovery with RIP as the criteria, and it means we need to consider how many sampled measurements are necessary to achieve the RIP. If we only focus on the dimensions of the problem (N is the dimension of signal, M is the number of measurements, and k is sparsity of signal), we have [57]

$$M \ \geq \ Ck \log(\frac{N}{k}), \tag{2.13}$$

where C is a small constant. Therefore, as the number of measurements is beyond the bound, original data can be recovered exactly in ideal condition.

Although RIP provides guarantees for the recovery of $k-$sparse signals, the verification that a general matrix D satisfies the property always demands a combinatorial search over all $\binom{N}{k}$ submatrices. In many cases,

[2] The name of sensing matrix and compressed matrix are not formal definitions in literature, the definition is just used for corresponding to the conception of compressed sensing.

we would like to verify the property of matrix D straightforwardly and easily for providing more concrete recovery guarantees. Thus, we introduce another criteria for evaluation, $\mu(D)$, named as coherence. [56] Coherence can be adopted to necessarily testify whether there is the sparsest solution from current acquired measurements.

The coherence of a matrix D, $\mu(D)$, is defined as the largest absolute value of inner product between any pair of columns d_i, d_j in D, written as

$$\mu(D) = \max_{1 \le i < j \le N} \frac{\mid (d_i, d_j) \mid}{\parallel d_i \parallel_2 \parallel d_j \parallel_2}. \tag{2.14}$$

In addition, the lower bound of $\mu(D)$ is Welch bound [58], which achieves at $\sqrt{\frac{N-m}{m(N-1)}}$ as each inner product between any pair is same. Please note that as $N \gg m$, the lower bound is roughly $\mu(D) \geq 1/\sqrt{m}$.

Next, we give the relation between μ and sparsity k so that the exact recovery of signal y is guaranteed. Actually, the core of recovery in compressed sensing is the pursuit of unique sparse solution θ from the measurements Ψy. From [59], it is stated if

$$k < \frac{1}{2}\left\{1 + \frac{1}{\mu(D)}\right\}, \tag{2.15}$$

then for each measurement $z \in R^m$, there exists at most one solution $\theta \in \Sigma_k$ such that $z = D\theta$, and θ is the sparsest solution. Therefore, the coherence $\mu(D)$ ensures that the unique sparse solution is found, and hence the exact recovery is obtained.

So far, we have basically introduced the background of this research, and next the related works are presented.

2.2 Related Works

The whole framework of the system is illustrated in **Figure 2.8**, which is basically composed of compressively acquisition, sparse solution estimation and computational reconstruction.

Thus, based on the framework, the related works in this section are discussed in aspects of reconstruction and acquisition. We firstly mention

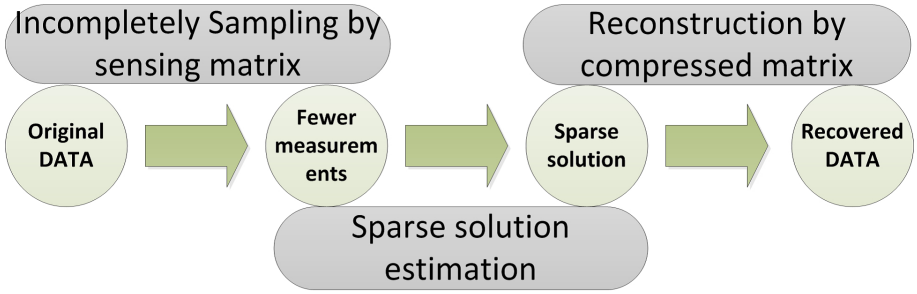


Figure 2.8: The whole framework of application of compressed sensing in ray space

the computational reconstruction of original signal from incompletely sampled measurements. Secondly, the related works on system for ray space acquisition are discussed.

2.2.1 l_p Norm Reconstruction

The estimation of sparse solution θ is the essential part of reconstruction, and it relates to l_p norm optimization. In fact, l_p norm is defined as

$$\| X \|_{l_p} = (\sum_{i=1}^{N} x_i^p)^{\frac{1}{p}}, \ p > 0. \tag{2.16}$$

Particularly, as $p = 0$, $\| X \|_{l_0}$ norm is defined as number of non-zero elements in X. The geometric property of l_p norm minimization is illustrated in **Figure 2.9** as $p = 2$, $p = 1$, and $p = 0.5$. As we can see, as $p \leq 1$, the optimization promotes sparse solution. Therefore, the reconstruction approaches are focused on l_p optimization as $p \leq 1$, and they can be roughly categorized into three parts including greedy algorithm, l_1 minimization, and iterative-shrinkage algorithm.

Greedy Pursuit: This type of method is usually dedicated to solving l_0 norm optimization, and the optimization formulation is written as

$$\hat{\theta} = \arg\min_{\theta} \| \theta \|_{l_0}, \ \text{subject to } y = \Psi\Phi\theta, \tag{2.17}$$

Greedy pursuit method, such as Matching Pursuits (MP) [60] and Orthogonal Matching Pursuit (OMP) [61], can provide a great flexible pro-

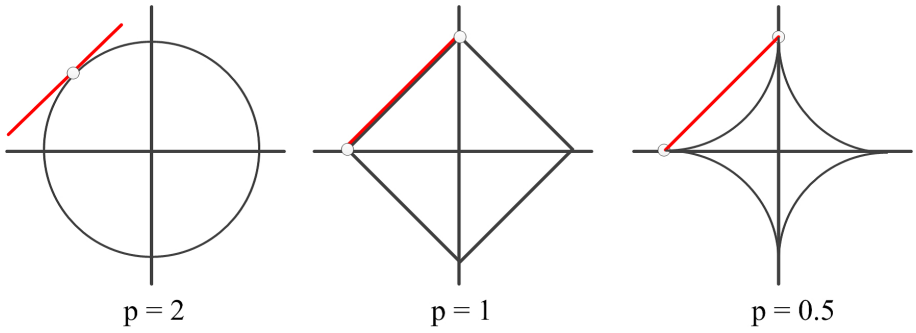


Figure 2.9: The geometric properties of l_p norm optimization

cess in estimating spares solution. Initially, greedy pursuit algorithm was adopted in signal decomposition. Similarly, $\Psi\Phi$ is assumed as one given dictionary and y is assumed as the given signal, thus the sparse solution is obtained if the measurement y is represented as linear combination of elements in $\Psi\Phi$ with the corresponding coefficients. Besides, the energy of signal can also be concentrated on the smallest number of elements since the largest inner product between residual and elements in dictionary is chosen in each iteration.

The basic procedure of greedy pursuit is illustrated as follow. If the signal is defined as f, decomposition space is named as space $\mathcal{H}$, and the element in the overcomplete dictionary D is called atom, named as d_i. Besides, the d_i is a unitary vector with the l_2 norm of 1. Thus, the first decomposition of f can be represented as

$$f = \langle f, d_0 \rangle d_0 + R^0 f, \tag{2.18}$$

where, d_0 is the optimal vector which can obtain the largest inner product with signal f , and $R^0 f$ is the residual of signal f. Clearly, if the iteration continues, it can obtain

$$R^n f = \langle R^n f, d_n \rangle d_n + R^{n+1} f. \tag{2.19}$$

Thus, another representation can be obtained after N iterations that

$$f = \sum_{i=0}^{N-1} \langle R^i f, d_i \rangle d_i + R^N f. \tag{2.20}$$

Besides, considering the energy, we have

$$\| f \|^2 = \sum_{i=0}^{N-1} | \langle R^i f, d_i \rangle |^2 + \| R^N f \|^2 \tag{2.21}$$

to guarantee the energy conservation.

Among the family of greedy pursuit algorithms, MP is one of the simplest methods. By using this method, the final residual signal will converge to certain predefined threshold ε. However, if the convergent speed is quite slow, it will lose the advantage of sparse representation. In contrast, OMP (Orthogonal Matching Pursuit) is an optimal approximation with respect to the selected subset of dictionary, because it ensures that the residual of signal is orthogonal to all the previous selected atoms in each iteration. The main difference between MP and OMP is that there is an update of coefficients in order to always keep the least error, rather than only choosing the largest inner product as the coefficient. The specific procedure of orthogonal matching pursuit is illustrated in Algorithm-1, and this algorithm will be frequently adopted in following chapters.

Furthermore, there are also other brothers for matching pursuit (MP) algorithms, such as CoSaMP [62], weak MP [63], Kernel MP [64], Adaptive MP [65], and etc. We will not go further in this direction which has been another hot topic in machine learning and sparse coding.

l_1 **Minimization**: Since l_0 norm optimization is not a convex optimization, it is always relaxed as l_1 norm optimization. Thus, the problem can be transferred to linear programming problem [66], which can be solved effectively and efficiently. The optimization process is modified as,

$$\hat{\theta} = \arg\min_{\theta} \| \theta \|_{l_1}, \text{ subject to } y = \Psi\Phi\theta. \tag{2.22}$$

Assume that the unknown solution θ is written as $\theta = u - v$, where $u, v \in R^n$ are both non-negative vectors such that u takes all the positive entries in θ and leaves the rest as zero. Similarly, v takes all the negative entries in θ while the rest are null. Due to this replacement, there is a concatenated vector, represented as $\mathbf{z} = [\mathbf{u^T}, \mathbf{v^T}]^\mathbf{T} \in R^{2n}$, and it is easy to see that $\| \theta \|_{l_1} = \mathbf{1^T}(\mathbf{u} + \mathbf{v}) = \mathbf{1^T z}$, and $\Psi\Phi\theta = \Psi\Phi(\mathbf{u} - \mathbf{v}) = [\mathbf{\Psi\Phi}, -\mathbf{\Psi\Phi}]\mathbf{z}$.

Algorithm-1 OMP

Input: $f_{rec} = 0, R^0 f = f, A = \{\emptyset\}, n = 0,$
$Ite_{num} = K$, D= {all the candidate atoms}

Repeat:

1: Compute inner product between candidate atoms and residual signal $\{\langle R^n f, d_i\rangle; d_i \in D\}$;

2: Find the largest inner product with the corresponding atom d_n, $|\langle R^n f, d_n\rangle| \geq \alpha \sup_i \langle R^n f, d_i\rangle, 0 < \alpha < 1$;

3: Record the atom and inner product in the set of A and C respectively;
$A_{n+1} = A_n \bigcup d_n, C_{n+1} = C_n \bigcup \langle R^n f, d_n\rangle,$

4: Find the optimal solution $\hat{C}_{n+1}$ for $arg\min_{\hat{C}_{n+1}} \| f - A\hat{C}_{n+1} \|_{l_2}$, update the coefficient set $C_{n+1} \rightarrow \hat{C}_{n+1}$ and residual signal $R^{n+1} f = f - A\hat{C}_{n+1}$,

Until: $n > K$ or $R^{n+1} f \leq \varepsilon$

Output: $\hat{C}, A$

Thus, the l_1 optimization problem is re-written as

$$\hat{\theta} = \arg\min_{z} \mathbf{1^T z}, \text{ subject to } y = [\Psi\Phi, -\Psi\Phi]z, \text{and } z \geq 0. \tag{2.23}$$

In order to keep the equivalence between l_1 norm optimization and linear programming, the supports of $\mathbf{u}$ and $\mathbf{v}$ can not be overlapped. If there is one solution, and the k-th entry in both $\mathbf{u}$ and $\mathbf{v}$ are non-zero. Without loss of generality, if we assume $\mathbf{u}_k > \mathbf{v}_k$, and take the replacement as $\mathbf{u}'_k = \mathbf{u}_k - \mathbf{v}_k$ and $\mathbf{v}'_k = 0$, we can satisfy both the positivity and the linear constraints and also eliminate the penalty by $\mathbf{u}_k - \mathbf{v}_k > 0$. Thus, there is no overlap between $\mathbf{u}$ and $\mathbf{v}$, and the problem in Eq. (2.23) is equivalent to a classical linear programming problem. There are numerous efficient methods for solving linear programming, such as Simplex algorithm [67], Interior point method [68], or Column generation [69], and fortunately, there are also several effective toolboxes available for solving the problem such as l_1-magic [70], CVX and L1-LS [71], SparseLab [72], GPSR [73], SparCo [74], and etc.

Iterative-Shrinkage Algorithms: The drawback of l_1 norm optimization is that it is only semi-convex. Besides, there is no close-form solution by adopting methods mentioned above. Alternatively, another simple strategy to attack the l_1 or l_p norm optimization is the Iterative-Reweighted-Least-Squares (IRLS) algorithm [75]. Different from the previous objective function, we would like to eliminate constraint in optimization and add a parameter λ to balance the penalty and constraint. The unconstrained objective function of Eq. (2.22) is written as

$$\hat{\theta} = \arg\min_{\theta} \| y - \Psi\Phi\theta \|_{l_2}^2 + \lambda \| \theta \|_{l_1}. \tag{2.24}$$

Then, we set $\mathbf{\Theta} = \mathbf{diag}(| \theta |)$, thus we have $\| \theta \|_{l_1} \equiv \theta^T \Theta^{-1} \theta$. By rewriting l_1 norm of θ, we may regard the l_1 norm as an (adaptively-weighted) version of the squared l_2 norm. Providing the estimated solution in current iteration is θ_{k-1}, we set $\mathbf{\Theta}_{k-1} = diag(| \theta_{k-1} |)$ and attempt to solve

$$\hat{\theta} = \arg\min_{\theta} \| y - \Psi\Phi\theta \|_{l_2}^2 + \lambda \theta^T \Theta_{k-1}^{-1} \theta. \tag{2.25}$$

After getting the approximated solution θ_k, the diagonal matrix Θ_k is generated and the iteration continues.

However, it is found that the IRLS algorithm is inefficient to get the final solution, and it either requires too many iterations or shows no tendency of convergence. This is especially the case for high dimensional problems, such as image processing or ray space processing. In recent years, a new family of numerical algorithms has been gradually built, addressing the optimization problem above efficiently, and the family is named as Iterative-Shrinkage Algorithms. Roughly speaking, a shrinkage factor is constructed by slightly modifying the optimization process so that the convergence is guaranteed. We will stop discussing specific algorithms because it has beyond the scope of the thesis. Please refer [59] to find other reviews and details.

2.2.2 Acquisition Systems of Ray Space

Next, let us go back to the field of ray space to discuss the acquisition system. In fact, ray space acquisition has always been a hot researching topic, and there are numerous of related works that have paved a great way for our research. Roughly, the acquisition systems can be divided into several types as follows.

Lenslet (Microlens): This method was initially proposed by Lippmann [4] and it was quite straightforward. Lenslet arrays or pinholes are placed in front of a film sensor, so that the angular images can be diverted to avoid integration on the final sensor. In recent years, this lenslet-based system was integrated into digital camera [43], and corresponding consumer product is now widely available. One example for this system is illustrated in **Figure 2.10**. However, an obvious drawback for this design is the low spatial resolution of each angular image. In other words, the total resolution of image sensor is fixed and shared by both of spatial resolution and angular resolution, thus a permanent trade-off of resolution has to be made before the acquisition.

Camera array: In order to enhance the spatial resolution while preserve the angular resolution. The method of camera array is an alternative.

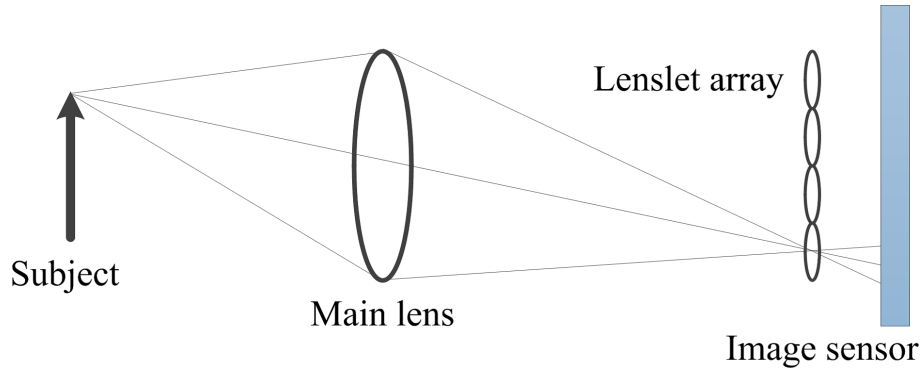


Figure 2.10: The illustration of ray space acquisition by lenslet system

Besides, another similar system is to resort a movable camera that travels back and forth to capture a group of images from different perspectives. In paper [76], the author presented a camera array with 128 cameras, named as Stanford multi-camera array. Besides, paper [77] has established a distributed light field camera by using 64 cameras. 100-camera system has also been developed in paper [78]. One example for this design is shown in **Figure 2.11**. However, the huge system requires high cost in equipment, computation, storage, etc. Besides, the calibration and rectification have also to be carefully conducted.

Multi-focus images: Besides camera array, recently, a single camera can also be adopted in ray space acquisition with high resolution. For instance, multi-focus images obtained by a single camera can be used to regenerate all-in-focus images for ray space synthesis. In paper [79], a method based on linear filters was proposed to reconstruct all-in-focus images for dense ray space synthesis. The all-in-focus images were reconstructed through shifted pinhole images which were obtained by the linear combination of multi-focus images. Although the method can avoid suffering from limitation of resolution of ray space, the data amount is still a problem for transmission or storage. The paper [80] attempted to compress the enormous data and developed one scheme for 4D ray space compression. The method firstly demonstrated synthesis and reconstruction between 4D ray space and multi-focus images, and then con-

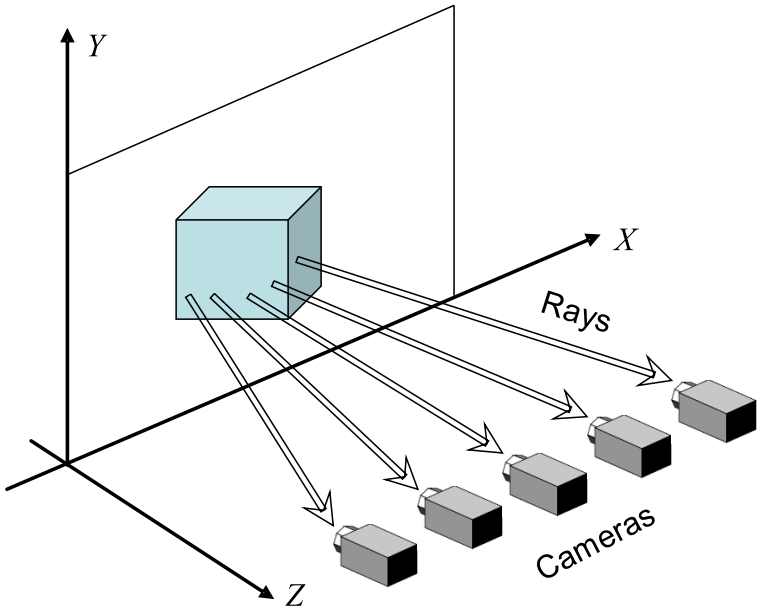


Figure 2.11: The illustration of ray space acquisition by camera array system

ventional methods were adopted to compress multi-focus images because multi-focus images that contained mostly low frequency components were easily compressed. Although this method is effective in compression, the huge amount of data still have to be obtained in the first place, thus the data problem in acquisition has not been solved.

Coded aperture/mask: Coded aperture imaging is proposed in order to collect more light in the situations that camera array system is too expensive to afford. This technique is based on the principle of pinhole imaging. The aperture plane is assumed as a combination of several "pinholes", and each "pinhole" corresponds to one angular image, but each angular image suffers from low signal noise ratio (SNR) as only one pinhole is open. Instead, a specially designed array of pinholes is adopted. The specifically designed mask (multiplexing pattern) attenuates several "pinholes" and opens the remaining "pinholes", so that angular images are linearly combined and the signal noise ratio (SNR) is also enhanced. If the multiplexing patterns are well designed, it provides a possibility to re-

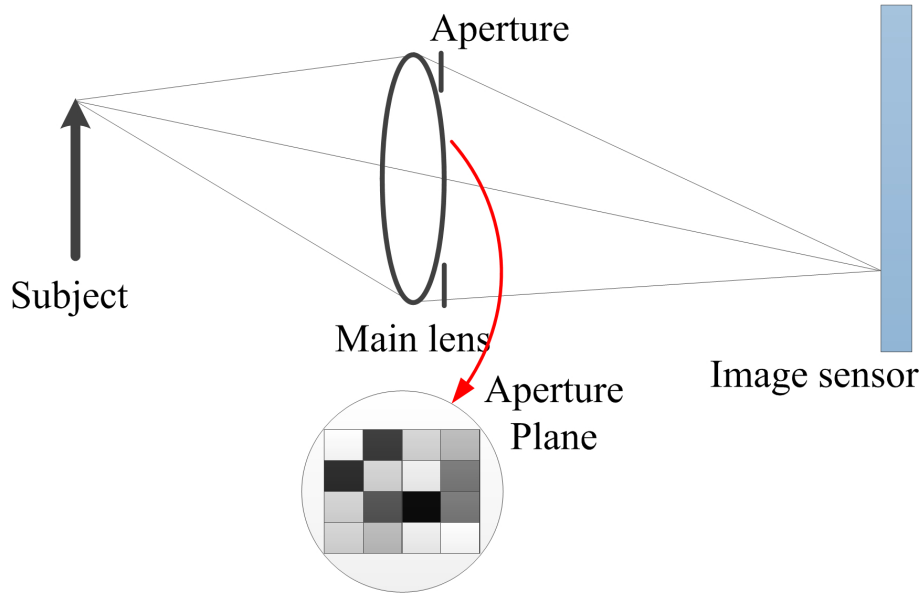


Figure 2.12: The illustration of ray space acquisition by coded aperture design

construct all of the angular images, and this technique is named as coded aperture. [81] One example of coded aperture is illustrated in **Figure 2.12**, and there are 16 "pinholes" in aperture plane, where the different gray levels represent the varied transmittances of each "pinhole".

After the coded aperture, all angular images will be integrated as one image in sensor. Therefore, the coded aperture pattern can be viewed as a multiplexing function. If this function is well designed, it can be equivalent to frequency analysis of ray space. More important thing is if a clear arrangement of optical elements is adopted, it is possible to re-bin the 4D rays and capture it using a 2D sensor [43, 82]. However, optical re-binning of rays is still based on microlens array system, and it requires a fixed and permanent tradeoff between spatial and angular resolution. Thus, the authors in [83] proposed another patterned mask to exploit the fact that light rays can be linearly combined together. Rather than sensing each 4D ray on the own pixel sensor, the design allowed sensing linearly independent weighted sums of rays. This method of mapping 4D

ray space to a 2D sensor array exploited heterodyning methods that were best described in the frequency domain. Besides, the designed mask was not limited in aperture plane, but it was placed in the optical path between aperture and sensor, to achieve Fourier domain remapping. Next, the authors in [84] took the extension, and interpreted "heterodyning" as a general theory (both of microlens-based and mask-based systems were applicable) of multiplexing the radiance in the frequency domain. By using the interpretation, the authors derived a mathematical theory of reconstructing the 4D spatial and angular information from the multiplexed 2D frequency representation. Authors in [85] developed a mathematical framework that generalized multiplexed imaging to all dimensions of the plenoptic function. This framework unified a wide variety of existing approaches to analyze and reconstruct multiplexed data in either spatial or frequency domain. Furthermore, other people in [86] tried to take 4D frequency analysis of ray space for the extension of depth of field by deconvolution, and the also showed an upper bound on the maximal power spectrum that could be achieved.

However, all the methods and proposals above require the equivalent acquisitions as the number of angular images in order to solve the inverse linear problem (deconvolution). Therefore, coded aperture solves the resolution and SNR problem and it provides an opportunity for frequency analysis as well. However, it shows no efficiency in the acquisition, and there is still distance to the compressive acquisition and analysis of ray space.

Compressive acquisition by coded aperture/mask: Next, we discuss the compressive acquisition of ray space. Actually, compressed sensing has been applied to video acquisition [87, 88], and image sensing [89]. Since a sequence of frames in video is similar as a group of angular images in ray space. Thus, compressive acquisition of a ray space is also possible by adopting the conception of compressed sensing. The authors in [90] simulated a compressive camera array for ray space acquisition. Furthermore, the light field super-resolution could be also regarded as a kind of compressive acquisition, because high resolution light field was expected to be recovered from low resolution measurements obtained

by mircolens-based lytro camera. [91] Moreover, view synthesis based on depth image [92] can also be viewed as a particular kind of compressive acquisition and reconstruction. [3]

Recently, researchers have started to explore compressive light field acquisition by employing only one sensor, and there have been several optical designs. In paper [93, 94], the authors proposed to employ coded aperture method in acquisition, however, the number of coded pattens are fewer than the pixel number of image patches. The computational reconstruction method was also developed for reconstruction. In addition, authors in [95] proposed and simulated a combination of coded masks (one was placed in aperture and the other one was mounted along the optical path) so that better reconstruction was obtained. However, the combination of two masks have reduced the light transmittance, resulting low SNR. In paper [96], the authors proposed to put an optimized mask between aperture and image sensor. For one thing, it increased the light efficiency, and for another thing, this kind of design also enhanced the randomness in acquisition, which was of significance in compressed sensing. In addition, it was reported from [96] that the reconstruction quality was acceptable only using single snapshot.

Therefore, all of the related works above provide us strong motivation and great foundation to proceed in this area for compressive acquisition and computational reconstruction of a ray space. In the following chapters, based on these related works, several novel approaches are proposed to achieve our targets.

2.3 Summary

In this section, we mainly discussed the background and related works of our research in this thesis. The background includes three parts, plenoptic function parameterization, computational photography and the applications, and the compressed sensing theory. We believe that it will be useful for readers to better understand the conception, methodology and appli-

[3] Only several images and corresponding depth maps are required, and the rest images can be synthesized by using geometric information

cation of our research.

The related works in this thesis are composed of reconstruction by l_p norm optimization, and system for acquisition of ray space. Rooted in the background mentioned above, our work is based on the accumulation of previous contributions from other researchers. Thus, it is necessary to have a review of related works. For one thing, it is better to distinguish our contributions from the previous technologies. For another thing, we give great appreciations and respects to the past contributors in this researching area, and it is their great and valuable works that gave us motivations, suggestions and inspirations. In the following chapters, we will discuss specific problems in the acquisition and reconstruction of a ray space.

Chapter 3

Compressive Acquisition and Reconstruction of Ray Space by Sparse Representation of Dictionary

In this chapter, since the general purpose of our work is to obtaining one ray space by compressive acquisition and computational reconstruction, we aim to conduct compressive acquisition of a ray space at first and develop two different types of dictionaries for enhancing the reconstruction quality of ray space.

We firstly establish the compressive acquisition model for ray space, and assume that the ray space has been compressively sampled. Next, we consider the methods to enhance the sparsity of ray space in dictionary domain, so that better reconstruction quality is expected. Since sparsity k is defined as the number of atoms in the representation of signal under the same level of distortion ε, we seek smaller k in different dictionaries. The similarity between the two types of dictionaries is that both of them are overcomplete so that we have more chance to select better atoms in sparse representation of signal. However, the difference of both dictionaries are also quite large. For designed dictionary, it has implicit structure and the dictionary is generated by tuning several parameters, thus the redundancy of this dictionary is large. While for learned dictionary, it has

no structure and it is obtained by learning method from other training data. In addition, it has to be presented by explicit matrix form.

The rest parts in this section are organized as follows, after the description of acquisition model, the sparsity analysis of ray space in orthogonal bases are discussed, and the sparseland model [59] is also introduced. Next, based on the sparseland model, we design one overcomplete dictionary whose number of atoms exceeds the dimension of dictionary for sparser representation of ray space so that the reconstruction quality of ray space is enhanced by adopting the carefully designed dictionary. In addition, we adopt dictionary learning to shape one overcomplete dictionary adaptively, so that the features of ray space can be represented more accurately and properly by the learned dictionary, resulting better reconstruction quality. Furthermore, we also present the relation between reconstruction quality and the number of acquired measurements.

3.1 Mathematical Model of Ray Space Acquisition

In this section, we mention the mathematical model for compressive acquisition of a ray space, and particularly the coded aperture is adopted as the example for description. In order to establish the model for compressive acquisition, we recheck the parameterization of a ray space, as illustrated in **Figure 3.1**. Plane (s,t) and (u,v) are image plane and sensor plane respectively, and a ray space is written as $L(u,v,s,t)$. Thus, a traditional photograph $i(u,v)$ is the projection of the ray space $L(u,v,s,t)$ along the image plane (angular dimension) (s,t), written as

$$i(u,v) = \int\int_{s,t} L(u,v,s,t)dsdt. \tag{3.1}$$

To be simple, we consider that both of 2D spatial and angular parameterizations are vectorized as one dimension, and the acquisition is rewritten as

$$i(\vec{u}) = \int_{\vec{s}} R(\vec{s},\vec{u})d\vec{s}, \tag{3.2}$$

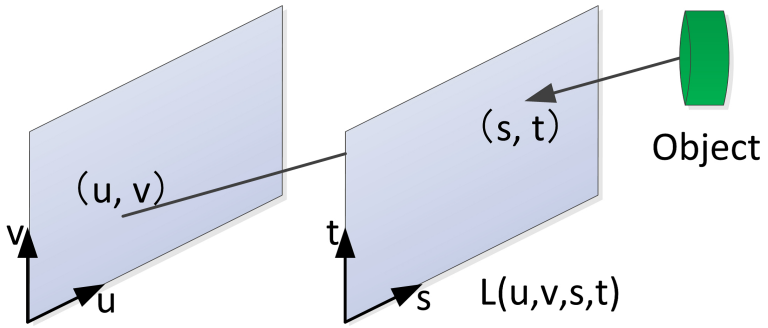


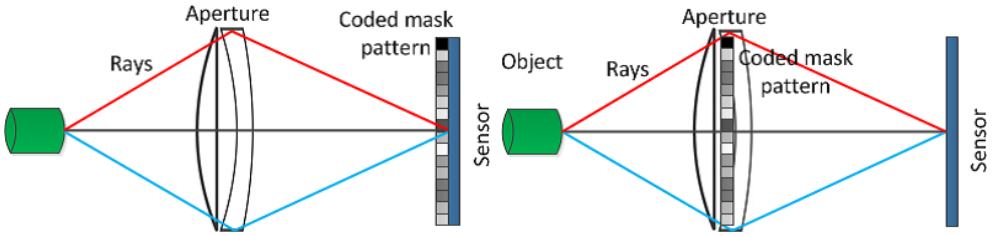
Figure 3.1: The 4D parameterization of ray space

where $\vec{u} = (u, v)$ and $\vec{s} = (s, t)$.

Assume that there is a coded attenuation mask $m(\vec{u}, \vec{s})$ inserted into the camera, and we discuss two cases for placing the coded mask at different positions. Firstly, as the mask is mounted at the sensor plane as shown in **Figure 3.2 (a)**, $m(\vec{u}, \vec{s}) = (const, \vec{s})$ which means all the angular images are averaged and only several pixels in the final integrated 2D spatial image are blocked. In other words, all the angular information has been lost. Secondly, as the mask is mounted in the opposite side, aperture plane (or named coded aperture) as shown in **Figure 3.2 (b)**, those angular images are linearly combined with various weights, represented as $i(\vec{s}) = \int\int_{\vec{u}} m(\vec{u}, const) \cdot R(\vec{u}, \vec{s}) d\vec{u}$. Thus, it provides an opportunity to reconstruct all of angular images. The basic principle for this model comes that different regions of aperture can acquire the angular images from different perspectives. As shown in **Figure 3.3**, the parallax of different angular images can be perceived clearly from the bricks in the background.[1]

Next, we assume the aperture is divided into N blocks, and each block corresponds to one angular image. In other words, N angular images will be multiplexed together. Besides, the transmittance of each block is assumed as m_j, then the final acquisition on the sensor can be represented as $Z^j = \sum_{j=1}^{N} m_j X^j$. Assume that $X^j_{K\times 1}$ is one vectorized angular image, and K is the resolution of angular image X^j. Therefore, one acquisition of ray space $R_{NK\times 1}$ by matrix multiplication can be formulated as $Z_{K\times 1} =$

[1]The data set comes from other data base, and we will mention it in next chapter.



(a) mask at sensor (b) mask at aperture

Figure 3.2: Different position for mounting a coded attenuation mask

$M_{K\times NK}R_{NK\times 1}$, where $R_{NK\times 1}$ is a long concatenated vector, and $R_{NK\times 1} = [X^1_{K\times 1}{}^T, X^2_{K\times 1}{}^T, ..., X^N_{K\times 1}{}^T]^T$. If $[m_1, m_2, ..., m_N]$ is one vector, the matrix $M_{K\times NK}$, also named as sensing matrix, can be generated by Kronecker product as $M_{K\times NK} = [m_1, m_2, ..., m_N] \otimes I_{K\times K}$, where $I_{K\times K}$ is an identity matrix, and $[m_1, m_2, ..., m_N]$ corresponds to the design of coded aperture. The illustration of one acquisition is shown in **Figure 3.4**. Clearly we can see, the pixels which are located in the same position of different angular images are multiplexed together by coded pattern, which means the relation of ray space is exploited. Thus, if the number of multiplex is smaller than the number of total pixel in angular image, the compressive acquisition is achieved and reconstruction by computational method is also possible.

So far, we assume one ray space has been compressively acquired. In the following part, we mainly discuss the reconstruction of ray space, and we start from the sparse representation of ray space, which plays an important role in the reconstruction of ray space.

3.2 Sparse Representation of Ray Space

Traditionally, an orthogonal transform, such as Discrete Fourier Transform (DFT) or Discrete Cosine Transform (DCT) is always employed as signal analysis model to obtain efficient decomposition and representation of signal, and it can be modeled as $Y = \Phi_{n\times n}X$, where $\Phi_{n\times n}$ is an orthogonal basis, Y is signal and X represents the analysis coefficients. Normally, due

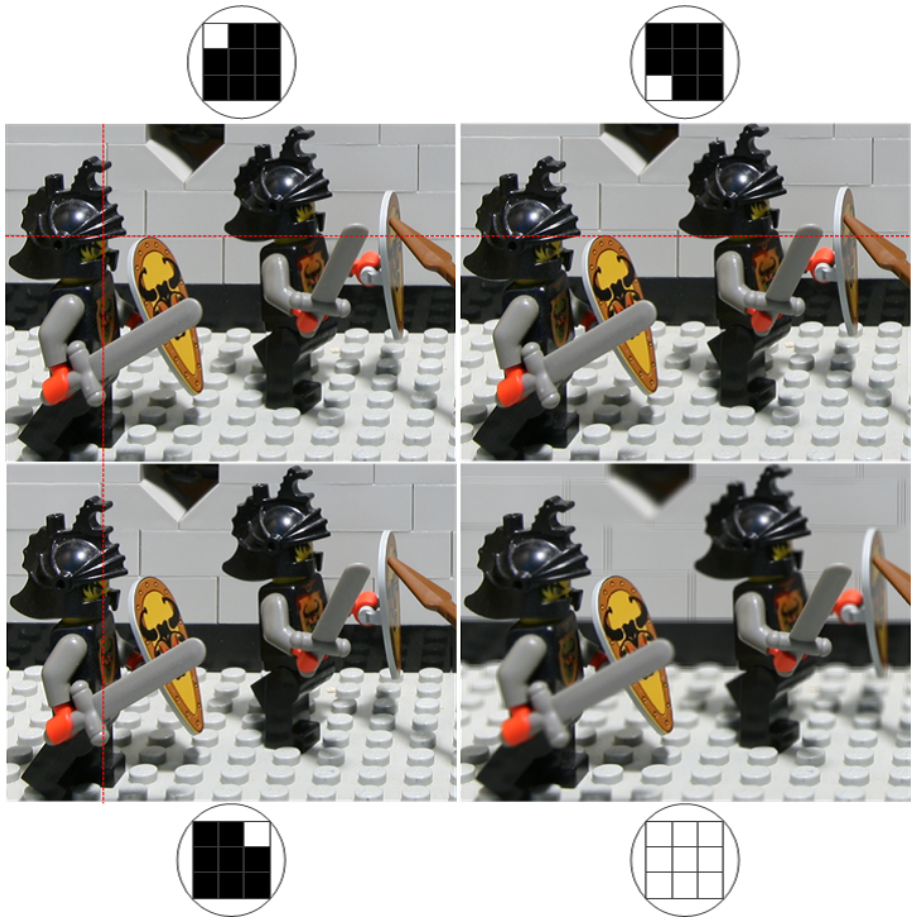


Figure 3.3: Different perspectives from different angular images

to the feature of natural signal, the coefficients in transformed domain shows the property that a few coefficients have quite large values while the rest coefficients have only small values. Thus, if there is a threshold τ for the coefficients X, where $X(i) = 0$ if $X(i) \leq \tau$, the signal can be approximately represented by a few coefficients with relatively small approximation error. This is traditional signal approximation, which plays key function in compression, pattern recognition, and etc.

In contrast, sparseland model [59] employs an overcomplete dictionary for signal approximation. Thus, the orthogonal basis $D_{n\times n}$ is extended

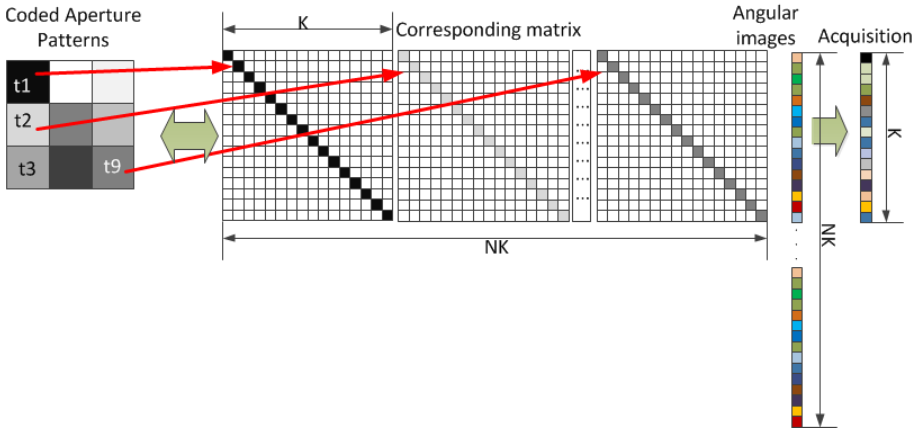


Figure 3.4: Acquisition of ray space by coded aperture

to be $D_{m\times n}, d_i \in R^n, m > n$, and each column in D is called an atom.[2] Therefore, the sparseland model can be described as $\mathcal{S}(D, X, k_0, \varepsilon)$. Given a signal Y as input, there will output the k_0-sparse representation X of Y under certain dictionary D with error ε. Therefore, in the following part, the task is to find a proper D that enables signal Y to be more sparsely represented.

Considering the target signal Y, we adopt epipolar plane image (EPI) as the processing unit. Since EPI shows clear structure as mentioned in last chapter, it is possible to have sparse representation. One epipolar plane image is illustrated in **Figure 3.5** as an example, and the EPI is generated by horizontally 'cutting' ray space "Fuzzy". [3] We can find easily that the EPI is composed of several straight lines with different slopes and intersections. In order to save the computation burden, the analysis and computation will be conducted based in terms of each patch of EPI. In the following content, we discuss the sparsity of EPI in several transform domains.

Fourier transform is regarded as the foundation of signal analysis and it is believed that most of natural signals in spatial domain can be analyzed

[2] In the rest of this chapter, we use D to represent an overcomplete dictionary

[3] "Fuzzy" ray space is from our own data base, the resolution of "Fuzzy" is $640 \times 480 \times 64$.

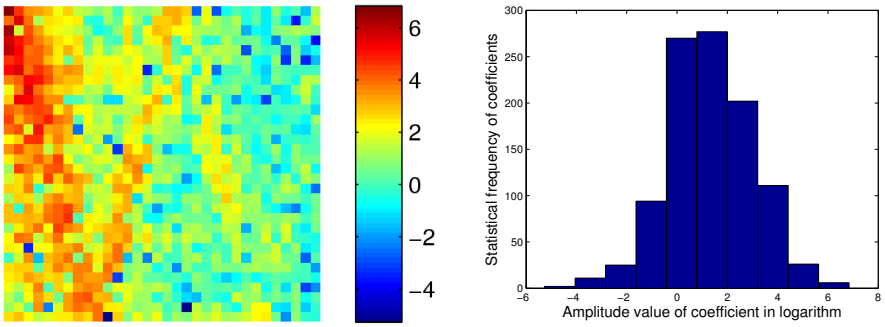
Figure 3.5: One exampling EPI in "Fuzzy" ray space

into a few coefficients in frequency domain. These kinds of analysis models are mainly attributed to the fast decay rate of k-term approximation, which means that it is possible to represent signal by the best k non-zeros elements from the transform coefficients.

Thus, we explore sparsity of EPI in frequency domain. The example of EPI is adopted and its spectrum in 2D-DCT domain is illustrated in **Figure.3.6**(a). Obviously, most parts of energy is concentrated in a small area (The red area in spectrum). To have a more explicit illustration, the histogram of DCT coefficient is also plotted in **Figure.3.6**(b), and the distribution of coefficient shows that EPI is sparse in frequency domain.

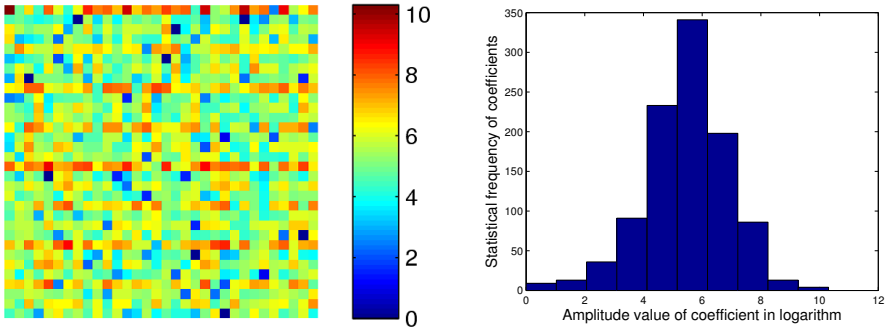
Besides, we also have a sparsity analysis of EPI in Welsh Transform domain, and Hadamard Matrix is employed for transform. Similar as frequency sparsity analysis, the spectrum and histogram of Discrete Hadamard Transform (DHT) coefficients, as shown in **Figure.3.7**, also imply that EPI is sparse in Welsh Transform domain as well. Even if the sparsity is only analyzed for one patch of EPI, it is believed that most of patches of EPI share the similar feature. Thus, we believe EPI is relatively sparse in transform domain, either by DCT or DHT, which implies that compressed sensing can be applied to EPI by using common orthonormal basis.

However, so far, the EPI is only treated as a normal signal, but the



(a) Spectrum in logarithm (b) Histogram of coefficient

Figure 3.6: The sparsity analysis of EPI in DCT domain.



(a) Spectrum in logarithm (b) Histogram of coefficient

Figure 3.7: The sparsity analysis of EPI in DHT domain.

special structure of EPI has not been taken into consideration. More suitable dictionary is supposed to be explored for sparser representation of EPI. Feature analysis is believed to obtain sparser representation, as long as another proper dictionary includes similar features of EPI.

3.3 Gabor Dictionary Design for Representation and Reconstruction of Ray Space

3.3.1 Gabor Dictionary Design

Since separable Gabor functions can be used to encode the motion compensated frames in video processing [97] due to the fact that Gabor function can grasp the feature of disparities in video sequence. Similarly, EPI actually reflected the disparity information between each FVI. Therefore, it is believed that the dictionary expanded by Gabor function will be an excellent candidate for providing the features of EPI. In addition, different from DCT or DHT whose transform matrix is an orthogonal square matrix, the transform matrix of Gabor dictionary is an overcomplete matrix. In other words, we extend the number of atoms in dictionary, and try to achieve sparser representation of EPI.

Normally, Gabor function is written as

$$g(x, y) = s(x, y)w(x, y), \tag{3.3}$$

where $s(x, y)$ is a complex sinusoid, named as carrier, and $w(x, y)$ is a 2D Gaussian-shaped function, named as envelope. To be specific, the carrier function is represented as

$$s(x, y) = \exp(j(2\pi(fx + hy)) + \varphi), \tag{3.4}$$

and the envelope function is written as

$$w(x, y) = K \exp\left\{-\pi\left\{\frac{(x-u)^2_{r_\theta}}{\sigma_1} + \frac{(y-v)^2_{r_\theta}}{\sigma_2}\right\}\right\}, \tag{3.5}$$

where (u, v) is the peak of the function, σ_1 and σ_2 are scaling parameters of the Gaussian function. Besides, the r_θ represents a rotation operation on x and y. From the structure analysis of EPI, it is clear that EPI is fluctuated in one direction and almost smooth in the other one. Therefore, in Gabor function, x is activated for varying while y is kept as constant. Besides, only the real part of carrier $s(x, y)$ is used. Then, 2D Gabor function can be simplified as

$$g(x, y) = \exp(-x^2 - y^2)\ \cos(2\pi(fx + hy) + \varphi). \tag{3.6}$$

Table 3.1: Parameter set of 2D Gabor dictionary.

$\sigma = 2^j, 0 \leq j \leq log_2 N$
$u = [0, N-1]$
$\theta = -\frac{6\pi}{12}, -\frac{5\pi}{12}, -\frac{4\pi}{12}, -\frac{3\pi}{12}, -\frac{2\pi}{12}, -\frac{\pi}{12}, 0$
$f = 0, 0.1, 0.2, ..., 1.0$
$\varphi = 0, \pi$

In addition, the Γ operator on x is given by

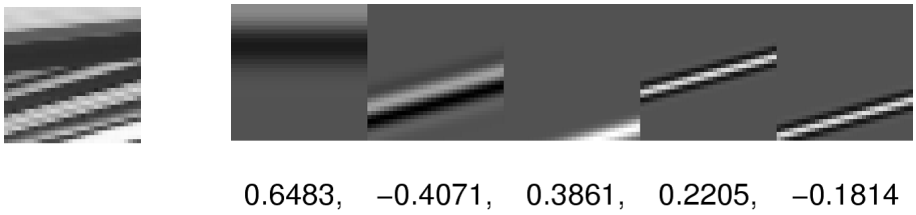
$$\Gamma g = g(\sigma^{-1}(x-u)_{r_\theta}), \tag{3.7}$$

and also the parameters f and φ are tunable. The parameter set can be represented as $P = \{\sigma, u, \theta, f, \varphi\}$, and numerous of atoms can be generated by tuning these parameters. The scope of parameters is given in **Table 3.1**, and N is the size of EPI patch.

Different from orthonormal basis, the sparsity exploration of EPI in Gabor dictionary is a non-linear transform. OMP is quite helpful in sparsity exploration since the output of OMP is exactly the selected atom and corresponding coefficient. Besides, the number of iteration in OMP corresponds to the sparsity. Thus, given the sparsity $s = 5$, for instance, one patch of EPI can be represented by 5 atoms with corresponding coefficients, as shown in **Figure.3.8**. Therefore, it is illustrated that Gabor dictionary is quite suitable for sparser representation of EPI.

Next, the sparsity comparison of EPI between common orthogonal basis and Gabor dictionary is conducted. In the comparison, EPI is divided into non-overlapped patches in size 8×8 [4]. The desired sparsity is arranged from $\{s = i | i = 1, 2, 3, ..., 10\}$. As for common orthogonal basis,

[4]Actually, each patch of EPI is unfolded as a column vector and transform is conducted by multiplying the transform matrix



(a) EPI block (b) Representation by Gabor atoms

Figure 3.8: Sparsity comparisons between basis and Gabor dictionary.

DCT and DHT are selected and the largest s coefficients will be preserved for the reconstruction of EPI patch. While for Gabor dictionary, sparsity s is the iteration of OMP. Root Mean Square Error (RMSE) was adopted as the evaluation of approximation error and it was written as

$$e = \sqrt{\frac{1}{H \times V} \sum_{1 \le i \le H, 1 \le j \le V} (I_{i,j} - \hat{I}_{i,j})^2}, \tag{3.8}$$

where I and $\hat{I}$ are original EPI and sparsely represented EPI respectively with the resolution of $H \times V$. The testing images are EPI from "Fuzzy" and "Kuma" ray spaces, illustrated in **Figure.3.9**. The approximation error is compared by given the same sparsity and results are shown in **Figure.3.10**. From both of the graphs, Gabor dictionary could obtain less approximation error in the same sparsity and it directly shows EPI enjoys sparser representation in Gabor dictionary compared to common orthonormal basis. In addition, as for "Fuzzy" EPI, DHT performs better since the disparity of "Fuzzy" EPI is larger and DHT is more likely to represent this feature than DCT. While for "Kuma" EPI, both of orthonormal bases get almost same performance.

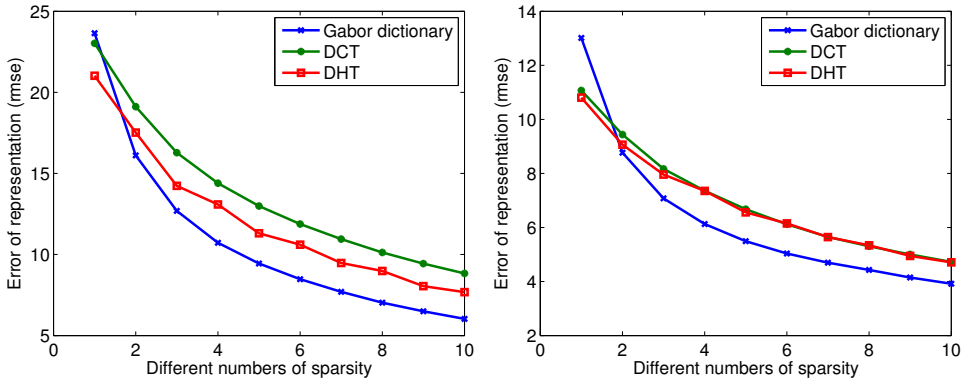
3.3.2 Experimental Results

In this part, we present experimental results for reconstructing the compressively sampled ray space. Our work focuses on the study how efficient ray space can be sensed, thus the experiments are based on synthesized

(a) EPI of "Fuzzy" ray space

(b) EPI of "Kuma" ray space

Figure 3.9: The examples of EPI in two ray spaces.



(a) Result for "Fuzzy" EPI (b) Result for "Kuma" EPI

Figure 3.10: Sparsity comparisons between basis and Gabor dictionary.

data, which means we have captured all the data at first as the ground truth. The framework of this experiment is illustrated in **Figure.3.11**. The ground truth of ray space and EPI are firstly obtained, and compressed sensing procedure is simulated. In the simulation, the measurements of EPI are obtained by adopting random sensing matrix. Next, the EPI and ray space are recovered by sparsity promotion algorithm. Finally, comparisons are conducted between the reconstructed ray space and the ground truth of ray space.

Before the experiment, it is necessary to conduct parameter configuration. Two testing ray spaces, named as "Fuzzy" and "Kuma" were em-

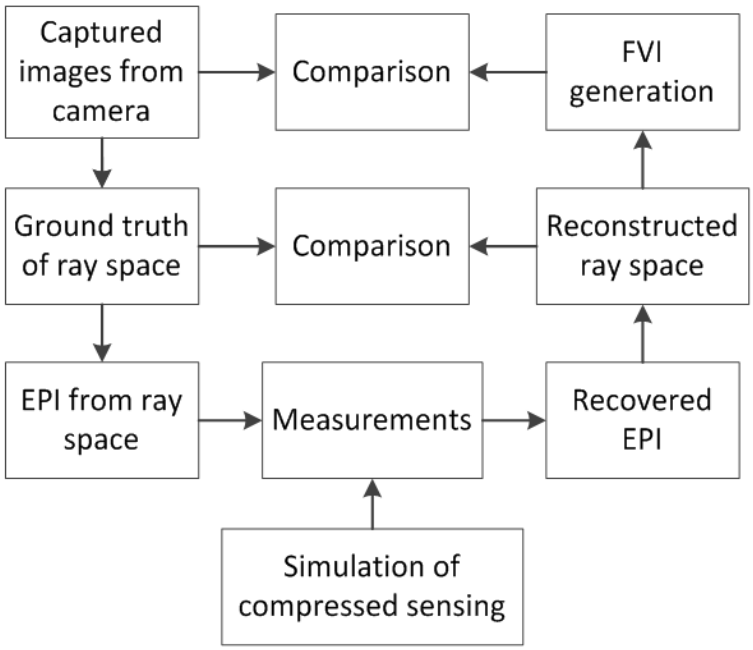


Figure 3.11: The whole framework of experiment.

ployed for this experiments. 64 cameras were equipped in linear arrangement for capturing images to generate "Fuzzy" and "Kuma" ray space, and each camera output one image with the resolution of 640×480. The images from one viewpoint for both ray spaces are shown in **Figure.3.12**. The content in "Fuzzy" ray space is more complex than the content in "Kuma" ray space. All the 64 images were combined together to establish the ray space with the resolution of $640 \times 480 \times 64$. Next, EPI was generated by cutting the ray space horizontally, and the resolution of EPI was 640×64, indicating that there were 480 EPI in the captured ray space. Two EPI samples from two ray spaces are shown in **Figure.3.9**. In the simulation, EPI was divided into non-overlapped 8×8 patches to avoid large computation and each patch is unfolded to be a column vector with 64 elements. Two random sensing matrices were adopted respectively. One was Bernoulli sensing matrix (The elements are 1 or -1) and the other was Gaussian sensing matrix. The sensing ratio R is defined as the fraction between the number of measurements and the total pixel num-

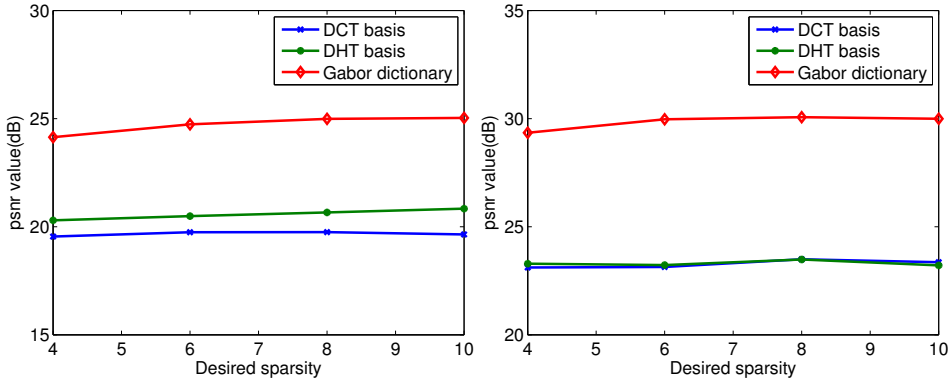
(a) FVI in "Fuzzy" ray space (b) FVI in "Kuma" ray space

Figure 3.12: FVIs in two ray spaces.

ber of ray space. For the recovery method, Orthogonal Matching Pursuit (OMP) was adopted for the estimation of sparse solution. Since sparsity of EPI was not available, we tested several desired sparsities as $s = 4$, $s = 6$, $s = 8$, and $s = 10$. Furthermore, in the selection of sensing ratio R, several sensing ratios were also tested, which were from 20% to 60% with the increment of 10%. It is true that reconstruction quality will be higher as the sensing ratio increases. However, if the sensing ratio is too high, closing to 100%, there is no meaning by adopting compressed sensing in acquisition.

In the first place, we test the effect of sparsity of EPI on the final recover quality by using a group of desired sparsities of EPI. In this test, the sensing ratio is fixed as $R = 50\%$. Besides, Bernoulli sensing matrix and Gaussian sensing matrix are adopted for "Fuzzy" and "Kuma" EPIs, respectively. Each test is repeated for 10 times and the averaged values are obtained and are illustrated in **Figure.3.13**. From the graphs, as the desired sparsity of EPI increases, especially over $s = 6$, the quality of recovery almost remains constant. Therefore, in the rest testing, we choose the desired sparsity of EPI as $s = 6$ for simplicity. In addition, Gabor dictionary achieves better results than the rest two at the same sparsity, which also means Gabor dictionary can exploit sparser representation of EPI than the rest two orthogonal bases.

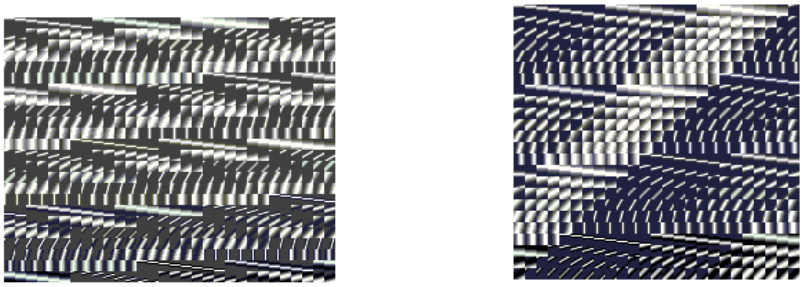
Next, we check the reconstruction performance of compressed sensing



(a)"Fuzzy" EPI (Bernoulli matrix)(b)"Kuma" EPI (Gaussian matrix)

Figure 3.13: The recovery performance by using different desired sparsities (Sensing ratio $R = 50\%$).

and the associated artifacts with different sensing ratios R. The experiment is conducted based on EPI, and different numbers of measurement are provided for testing the performances. The orthogonal bases, including DCT and DHT, and the designed 2D Gabor dictionary are used in experiment. As for 2D Gabor dictionary, since it is a redundant dictionary, in order to reduce the size of dictionary, statistics is also conducted in advance for removing the atoms which have never been used in sparse representation. Then it outputs a carefully designed dictionary as shown in **Figure.3.14**. Since the experiment is conducted in terms of EPI patch and matrix unfold is also conducted, the size of DCT and DHT is 64×64. Similar as the previous test, two random sensing matrices, Bernoulli and Gaussian, are selected. The testing EPI is randomly chosen from each ray space and each test is repeated for 10 times individually. The averaged performances (PSNR value) of "Fuzzy" EPI and "Kuma" EPI are shown in **Figure.3.15**. From the graphs, clearly the construction artifact is reduced (or PSNR value is increased) as the sensing ratio increases for both orthogonal bases and Gabor dictionary. In addition, since both of orthogonal bases are quite general, the structure of EPI can not be grasped well by either of them. Conversely, Gabor dictionary could grasp the feature



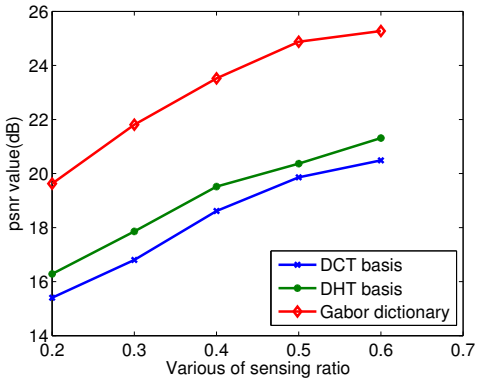
(a) Gabor dictionary for "Fuzzy" (b) Gabor dictionary for "Kuma"

Figure 3.14: The designed Gabor dictionaries for two ray spaces.

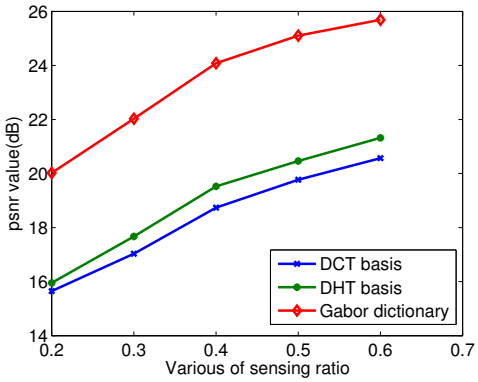
of EPI for effective sparse representation. Thus, 2D Gabor dictionary can achieve about $4-6$ dB increase compared to the other candidates. In the comparison between two bases, DHT performs a little better than DCT which corresponds to the result in sparsity comparison in **Figure.3.10**. What has to be mentioned is that if the sensing ratio continues increasing, better performance can be achieved. Furthermore, if more advanced optimization tool is adopted instead of OMP, the performance can be also enhanced.

Moreover, for subjective test, recovered EPI from measurements by three compressed matrices are presented in **Figure.3.16**. The sensing ratio is set as $R = 50\%$, and Gaussian random sensing matrix is adopted as sensing matrix. The left column shows recovered "Fuzzy" EPI by using different compressed matrices in compressed sensing, while the right column shows the performance of "Kuma" EPI. Subjective test gives a clear comparison to show that Gabor dictionary achieves the best reconstruction result.

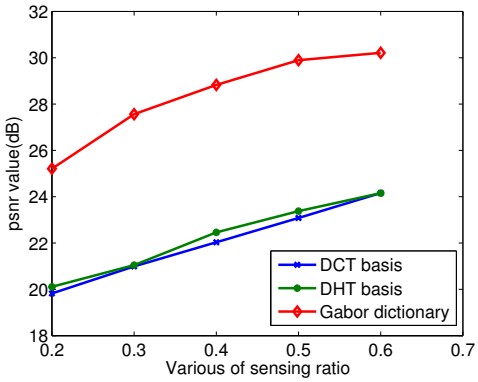
Finally, as all the EPIs are recovered, the whole ray space can be reconstructed. In this simulation, Gabor dictionary is adopted as compressed matrix and Gaussian matrix is adopted as sensing matrix respectively. The desired sparsity is set as $s = 6$. Since the reconstructed ray space could produce 64 viewpoint images, we selected No.26 as a virtual view-



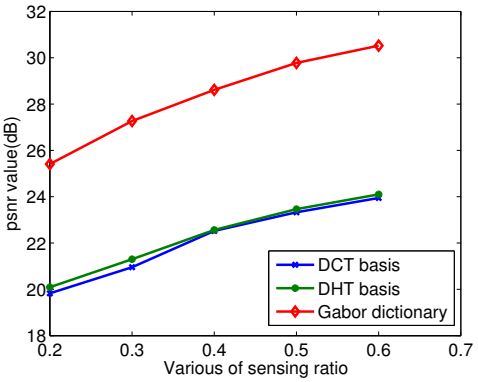
(a) 'Fuzzy' by Bernoulli Matrix



(b) 'Fuzzy' by Gaussian Matrix



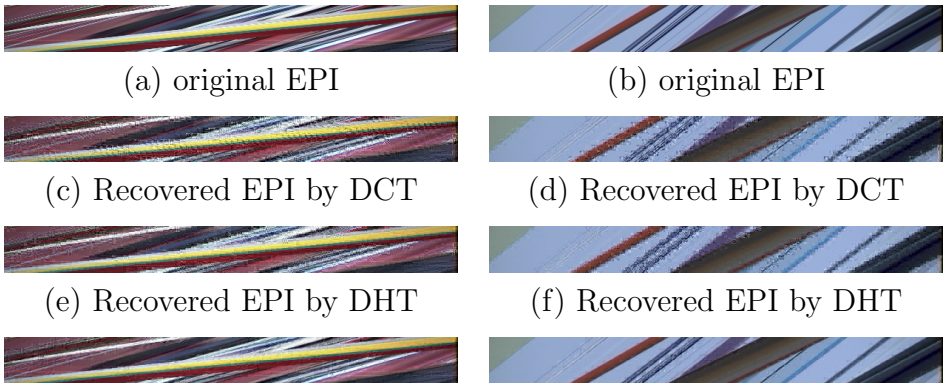
(c) 'Kuma' by Bernoulli Matrix



(d) 'Kuma' by Gaussian Matrix

Figure 3.15: Performances of recover among two sensing matrices by given different sensing ratios (Desired sparsity $s = 6$).

point. The generated FVIs from reconstructed ray space by using different numbers of measurement are illustrated in **Figure.3.17** and **Figure.3.18**, respectively. The statistical properties of each generated FVI are also presented in **Table 3.2**. Clearly, as sensing ratio increases, both of the mean and standard deviation of pixel error are reduced. In addition, the PSNR values of each generated FVI are also illustrated. Please note that the content of "Fuzzy" ray space is very complex, and hence the quality of generated FVI is relatively low. In contrast, for "Kuma" ray space, the quality of generated FVI is much better. Furthermore, the subjective eval-



(a) original EPI (b) original EPI

(c) Recovered EPI by DCT (d) Recovered EPI by DCT

(e) Recovered EPI by DHT (f) Recovered EPI by DHT

(g) Recovered EPI by Gabor (h) Recovered EPI by Gabor

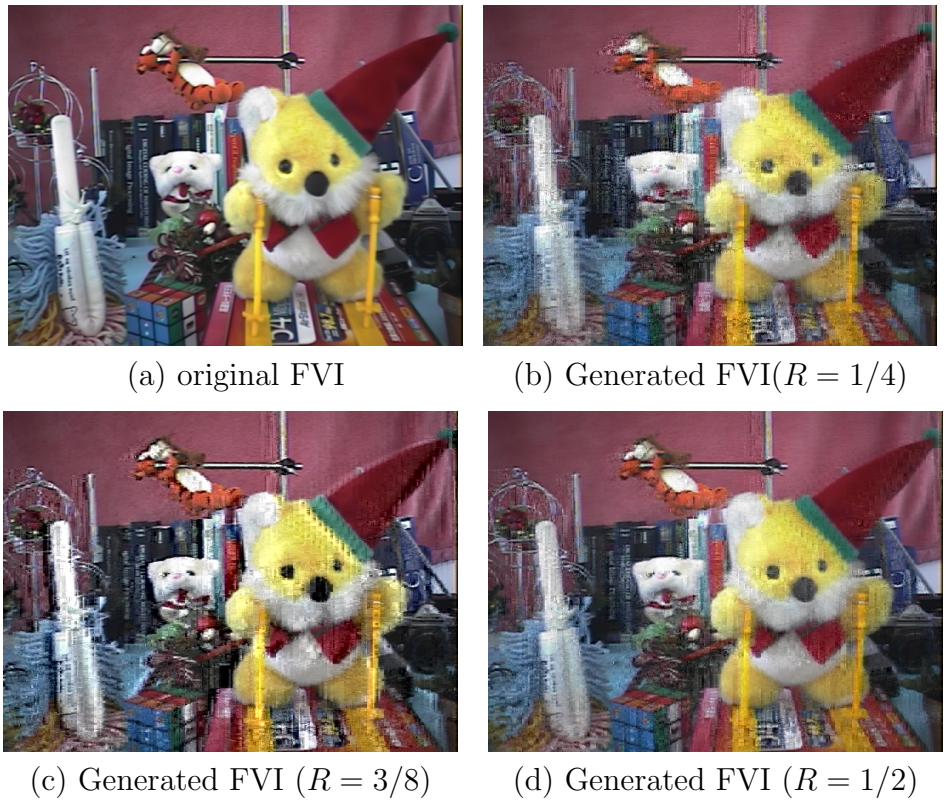
Figure 3.16: Recovered EPI by using different compressed matrices ($R = 50\%$ and Gaussian matrix is used for random sensing)

Table 3.2: Statistical property of generated FVI from different numbers of measurement

Sensing ratio	PSNR (dB)	Mean of error	Std of error
Fuzzy (1/4)	21.3517	14.2477	16.5329
Fuzzy (3/8)	23.7167	10.6972	12.7236
Fuzzy (1/2)	25.2532	9.1273	10.5201
Kuma (1/4)	26.4284	5.8540	10.6641
Kuma (3/8)	28.2430	4.7058	8.6778
Kuma (1/2)	30.0319	3.7144	7.1241

uation of generated FVI shows that the recovered result is quite acceptable by sensing a quarter of measurement. As the number of measurement increases to a half, the quality of FVI is improved greatly.

In conclusion, by taking specific analysis of sparsity of EPI and taking advantage of special feature of EPI, we explored a dictionary generated by Gabor function so that the special feature of EPI was grasped in this dictionary and sparser representation was achieved. It was shown in experimental results that compressed sensing method was applicable in the reconstruction of EPI and ray space, which led us to further design specific device for capturing natural ray space. Besides, it was also proved that



(a) original FVI (b) Generated FVI($R = 1/4$)

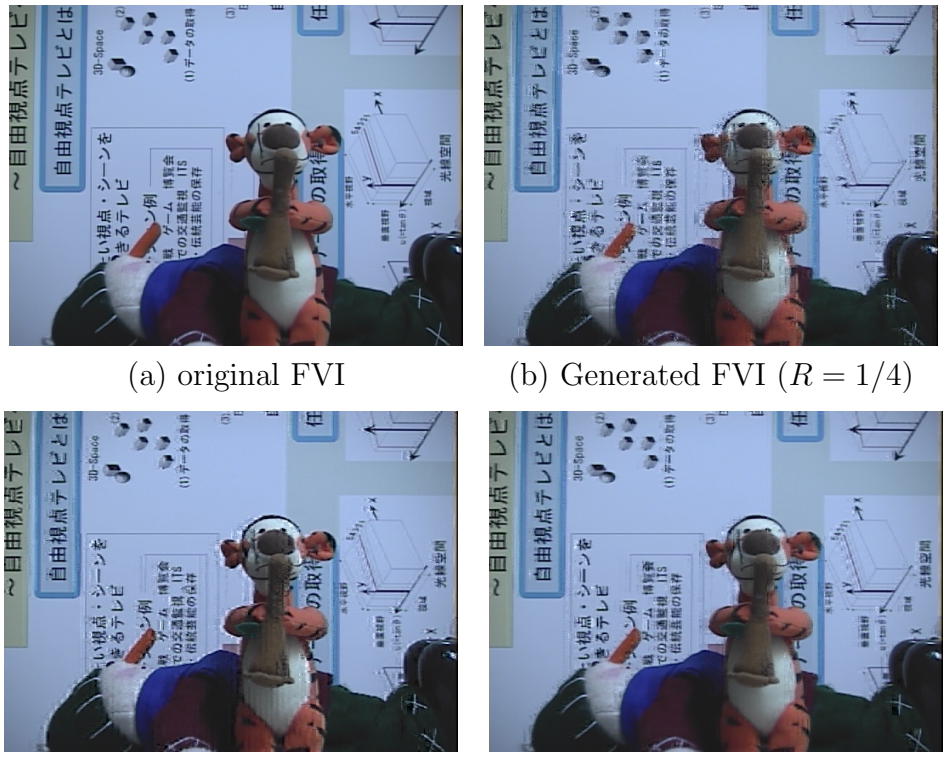
(c) Generated FVI ($R = 3/8$) (d) Generated FVI ($R = 1/2$)

Figure 3.17: "Fuzzy" FVI from reconstructed ray space by using different number of measurements.

the dictionary provided by Gabor function outperformed orthogonal basis, which also indicated us to explore other more powerful dictionary for achieving better reconstruction results. Therefore, in the next section, we will introduce dictionary learning method and integrate the well-learned dictionary in the compressive acquisition and reconstruction of ray space to expect better performance.

3.4 Dictionary Learning for Representation and Reconstruction of Ray Space

In this section, we extend the dictionary generation from designing to learning. The success of the designed Gabor dictionary in last section



(a) original FVI (b) Generated FVI ($R = 1/4$)

(c) Generated FVI ($R = 3/8$) (d) Generated FVI ($R = 1/2$)

Figure 3.18: "Kuma" FVI from reconstructed ray space by using different number of measurements.

is attributed to a number of atoms for representing different features of EPI. However, huge dictionary size also brings great burden in the estimation of sparse solution, especially by greedy methods. In addition, the similarity between the feature of EPI and the feature in 2D Gabor dictionary determines good performance of dictionary. Therefore, if there is a dictionary that can be adaptively learned from a set of training data, the feature of data can be grasped effectively and, more importantly, the size of dictionary can also be controlled in a small number for efficient computation.

In the following parts of this section, we firstly discuss dictionary learning procedure for sparse representation. Next, we apply the learned dictionary to the reconstruction of ray space. Finally, the experimental results

show that the learned dictionary is better than orthogonal basis and designed 2D Gabor dictionary.

3.4.1 Dictionary Learning

The essence of sparse representation by dictionary learning [98–100] is that an observation of signal $Y_{N\times L} = [y_1, y_2, ..., y_L]$, where each column $y_i \in R^N$ is a signal, can be sparsely represented by a few of elements from a dictionary $D_{N\times M} = [d_1, d_2, ..., d_M], d_i \in R^N, M > N$ which can be adaptively learned from a set of training data. Each column d_i in dictionary $D_{N\times M}$ is one atom, and $X_{M\times L}$ is composed of the corresponding coefficients of atoms in dictionary, resulting

$$Y_{N\times L} = D_{N\times M} X_{M\times L}, \tag{3.9}$$

where $X_{M\times L}$ is the matrix whose column vectors are sparse.

The whole procedure of overcomplete dictionary learning consists of two parts. One is sparse coding and the other is dictionary update. In the sparse coding stage, the dictionary $D_{N\times M}$ is assumed to be known and the sparse representation $X_{M\times L}$ is obtained by adopting greedy pursuit algorithm, such as Orthogonal Matching Pursuit (OMP) [61], or other optimization methods with sparsity promotion. In the part of the dictionary update, the obtained sparse representation $X_{M\times L}$ is employed with the signal $Y_{N\times L}$ for updating the dictionary $D_{N\times M}$. Thus, dictionary is updated gradually by each iteration k. In the following, we will take EPI as an example to illustrate dictionary learning for sparse representation of ray space.

3.4.2 Sparse Representation of Ray Space by Learned Dictionary

It is assumed that one EPI is partitioned into L non-overlapped patches in size $\sqrt{N}\times\sqrt{N}$ and each patch is unfolded to one vector $y_{N\times 1}$to be stacked in a signal matrix $Y_{N\times L} = [y_1, y_2, ...y_L]$, where L is the number of signal. Thus, we would like to shape one dictionary, $D_{N\times M} = [d_1, d_2, ..., d_M], d_i \in$

R^N, where M is the number of atoms. Since the dictionary is overcomplete, we set $M > N$. To enforce sparsity, $Y_{N\times L}$ is supposed to be approximated by only few combinations of atoms in $D_{N\times M}$ with corresponding coefficients in $X_{M\times L} = [x_1, x_2, ..., x_L]$. It can be specifically formulated as

$$\begin{aligned} &\arg \min_{D_{N\times M}, X_{M\times L}} \|Y_{N\times L} - D_{N\times M}X_{M\times L}\|_F^2, \\ &\text{subject to } \|x_i\|_{l_0} \leq s, x_i \in R^M, 1 \leq i \leq L \end{aligned} \tag{3.10}$$

where $\| \bullet \|_F$ is frobenius norm and $\| \bullet \|_{l_0}$ is the l_0 quasi norm. In addition, s represents the number of non-zero elements in x_i and $s << M$.

Since there are two unknowns, $D_{N\times M}$ and $X_{M\times L}$, in Eq. (3.10), an alternate projection method is adopted. Firstly, $D_{N\times M}$ is fixed to obtain sparse representation $X_{M\times L}$, and it can be formulated as

$$\begin{aligned} &\arg \min_{x_i, 1\leq i\leq L} \| y_i - D_{N\times M}x_i \|_{l_2}, \\ &\qquad \text{subject to } \| x_i \|_{l_0} < s. \end{aligned} \tag{3.11}$$

It is a NP-hard problem to find the globally optimal solution for this problem with l_0 regularization. Thus, we adopt OMP to find the suboptimal solution $X_{M\times L}$. Next, in the stage of dictionary update, the optimization problem is written as

$$\begin{aligned} &\arg \min_{D_{N\times M}} f(D_{N\times M}) = \| Y_{N\times L} - D_{N\times M}X_{M\times L} \|_{l_2}^2, \\ &\qquad \text{subject to } d_i^T d_i = 1, 1 \leq i \leq M. \end{aligned} \tag{3.12}$$

Next, we take the derivative of f with respect to $D_{N\times M}$ and set it to be zero, and it is easy to obtain

$$\frac{\partial f}{\partial D_{N\times M}} = -2(Y_{N\times L} - D_{N\times M}X_{M\times L})X_{M\times L}^T = 0. \tag{3.13}$$

Thus, finally we have

$$D_{N\times M} = Y_{N\times L}X_{M\times L}^T(X_{M\times L}X_{M\times L}^T)^{-1}. \tag{3.14}$$

At the end of one iteration, each column of the dictionary is l_2 normalized. Therefore, after several iterations, the dictionary $D_{N\times M}$ is expected to be well shaped for sparse representation of $Y_{N\times L}$.

Obviously, the main purpose of dictionary learning is to shape the dictionary so that the approximation error of given signal is minimized, thus the dictionary can fully represent the feature of EPI. In addition, the size of dictionary, M, can be preset before learning, we can keep the dictionary size compact to reduce the computation cost.

The specific procedure of learning a dictionary D for sparse representation of EPI is illustrated in Algorithm-2.

Algorithm-2 Dictionary learning for sparse representation

Input: Training EPI $Y_{N\times L}$, Initial dictionary $D_{N\times M}$

Set:
Desired sparsity in sparse coding: k,
Iterations in learning: J,
Starting j: $j = 0$
Repeat: $j \leq J$
1: Compute sparse solution X of EPI Y in current given dictionary D by using OMP method and there are only k elements in each row of X.

2: Based on the obtained sparse solution X, the updated dictionary can be calculated by $D_{(j+1)} = YX_{(j)}^T(X_{(j)}X_{(j)}^T)^{-1}$.

3: Normalize each column of D to keep $d_i^T d_i = 1$.

Until: $j > J$

Output: dictionary D, sparse solution X

3.4.3 Reconstruction of Ray Space by Learned Dictionary

Next, the overcomplete dictionary mentioned in previous part is adopted in compressed sensing framework.

In the sensing process, an EPI $Y_{N\times L}$ is measured after it is projected by a sensing matrix, $\Psi_{P\times N} = [\psi_1, \psi_2, ...\psi_p]^T$, where $\psi_i \in R^N, 1 \leq i \leq P$. The projected measurements $Z_{P\times L} = [z_1, z_2, ...z_L]$, where $P < N$, can be formulated as $Z_{P\times L} = \Psi_{P\times N} Y_{N\times L}$. Thus, the number of samples that are

actually measured is reduced from NL to PL. Please note that L was the number of training data in the previous learning procedure, but here we refer to it as the number of EPI patches in the whole ray space which we want to obtain.

Next, in the recovery stage, we consider to recover y_i from z_i, and finally $Y_{N\times L}$. However, y_i is not directly recovered, but the sparse representation x_i is recovered in the first place. The approximated sparse representation $\hat{x}_i$ is explored from z_i and the merged matrix $(\Psi D)_{P\times M}$ by optimization with sparsity promotion as follows:

$$\hat{x}_i = \arg \min_{x_i, 1\leq i\leq L} \parallel z_i - (\Psi D)_{P\times M} x_i \parallel_{l_2}, \text{subject to } \parallel x_i \parallel_{l_0} \leq \epsilon \quad (3.15)$$

Similarly, in order to find the globally optimal solution for this problem with l_0 regularization, the exhaustive sweep has to be conducted through all the possible supports, $\binom{s}{N}$ and it is also a NP-hard problem. Thus, OMP is employed to search the sub-optimal solution. Then, from the recovered $\hat{x}_i$, the EPI $\hat{y}_i$ can be obtained from

$$\hat{y}_i = D_{N\times M} \hat{x}_i. \quad (3.16)$$

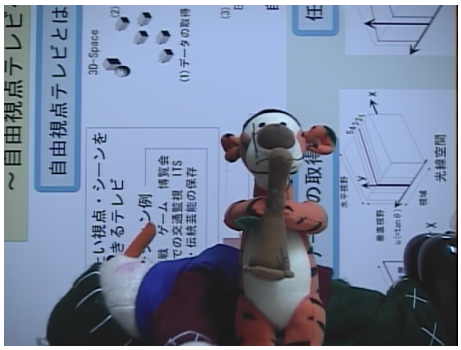
Therefore, the learned dictionary has been integrated into the reconstruction of compressively sampled ray space. In the next subsection, we will illustrate the experimental results.

3.4.4 Experimental Results

We firstly checked whether the proposed dictionary can have a sparser representation of EPIs. Secondly, we simulated the sensed measurements by using random sensing matrix, and then obtained the recovered EPIs by adopting OMP method. The recovered results by the proposed dictionary and other dictionary or basis were compared to show the advantage of the proposed dictionary.

Two data sets, named as "Fuzzy" and "Kuma" were employed throughout the experiments. There were 64 viewpoints closely arranged and the resolution of each viewpoint image was 640×480. One viewpoint images (No.25) from both data sets are shown in **Figure 3.19**. All the 64 images

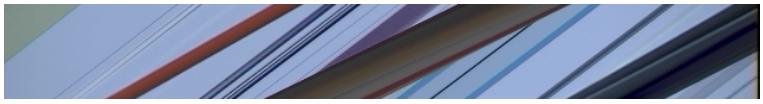
(a) FVI in “Fuzzy”



(b) FVI in “Kuma”

Figure 3.19: FVIs in two ray spaces

(a) EPI in “Fuzzy” ray space



(b) EPI in “Kuma” ray space

Figure 3.20: EPIs in two ray spaces

were aligned together for each data set to construct the ray space with the resolution of $640 \times 480 \times 64$. Next, EPIs were generated by ‘cutting’ the ray space horizontally, and the resolution of an EPI was 640×64, indicating that there were 480 EPIs in the constructed ray space. EPIs from both ray spaces are shown in **Figure 3.20**. It can be observed that the EPI from ’Fuzzy’, which included more objects with different depth, is more complex than the EPI from “Kuma”. Each EPI was divided into 8×8 patches and each patch was unfolded to be a 64×1 vector y_i. In addition, the learned dictionary by the method of optimal direction (MOD) [98] is illustrated in **Figure.3.21**.

Three candidates, orthogonal basis (DHT), 2D Gabor dictionary in [101], and the learned dictionary are compared in terms of sparsity. The

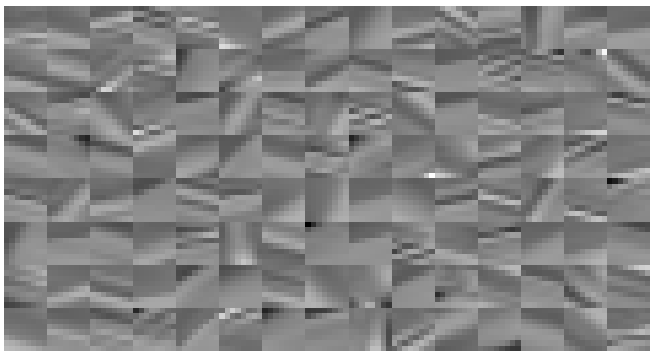
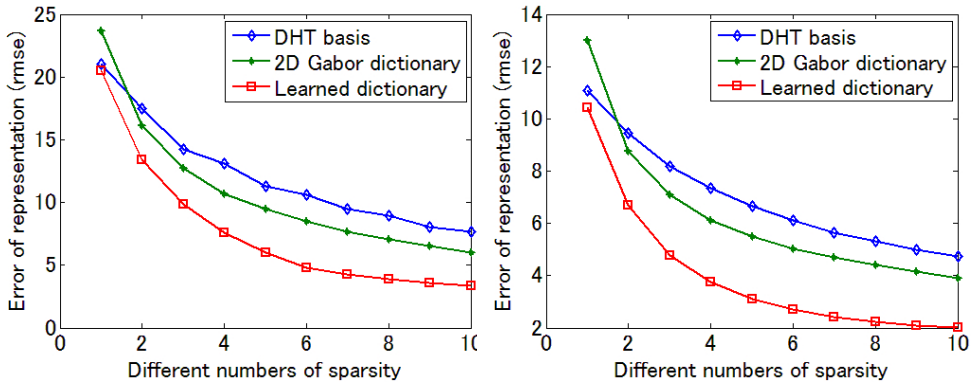


Figure 3.21: The illustration of dictionary by learning method

size of orthogonal basis is 64×64, and only the largest s coefficients in X are used for the approximation. The structured dictionary is highly redundant with the size of 64×740 for "Fuzzy"and 64×610 for "Kuma" ray spaces respectively. [5] The size of the proposed dictionary is only 64×128, which is much smaller than the structured dictionary, and it is learned from 7200 random-chosen EPI patches in 20 iterations. In **Figure 3.22**, the horizontal axis represents the desired sparsity s in sparse representation, while the vertical axis corresponds to the error which is evaluated by root mean square error (RMSE). Clearly, from the graphs, the learned dictionary gets the best performance for both ray spaces and achieves the least error for the same sparsity.

Next, the three candidates mentioned above are compared in the framework of reconstruction of ray space from compressive sampled measurements. A Gaussian random matrix is adopted as the sensing matrix for all

[5]The original size of structure dictionary was huge, beyond 3000, due to the large parameter space. To keep the dictionary size acceptable, we analyzed a statistics in sparse coding and kicked out the atoms which were rarely used in representation.

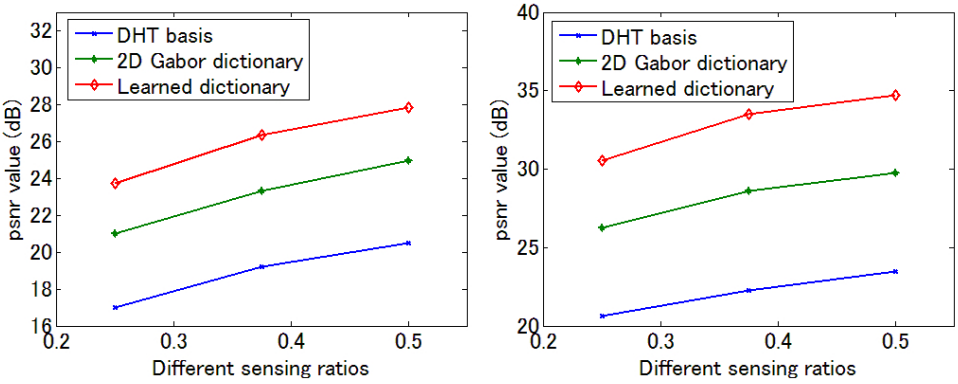


(a) Result for "Fuzzy" EPI (b) Result for "Kuma" EPI

Figure 3.22: Sparse representation of EPI in orthogonal basis, 2D Gabor dictionary and learned dictionary

the cases, and the sensing ratio R, which is the fraction between the number of sensed data and the number of original pixels, is set to 1/4, 3/8 and 1/2 respectively. The desired sparsity was set to $s = 6$ in the recovery of a randomly selected EPI, and the reconstruction errors (PSNR values) are compared among the three candidates. Since the sensing procedure was random, we repeat the sensing and recovery for 10 times and the averaged reconstruction errors are presented in **Figure 3.23**. From the graph, the learned dictionary can achieve 3-5 dB improvement in average compared to the 2D Gabor dictionary, and is much better than other orthogonal bases.

Since the dictionary adopted above is learned from parts of "Fuzzy" and "Kuma" ray spaces, it is better to adopt another dictionary that is learned from other ray spaces. Thus, we adopt other ray spaces from Stanford archives as training data set for dictionary learning. The comparison between the two dictionaries is illustrated in **Figure 3.24**. From the graph, we can see the dictionary learned from other data set performs worse than the dictionary learned from the same data, but it still has advantage over other candidate method, such as Gabor dictionary and orthogonal basis (DHT). Thus we claim that the dictionary learning method



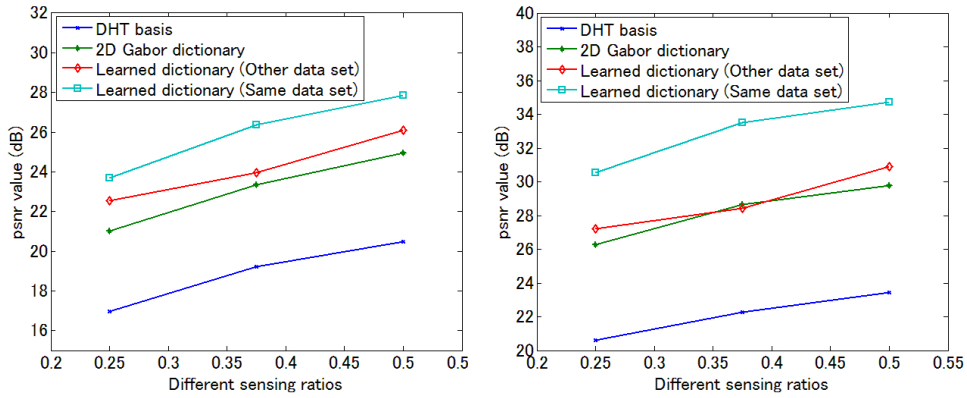
(a) Result for "Fuzzy" EPI (b) Result for "Kuma" EPI

Figure 3.23: Recovery results of two EPIs by using different sensing ratios

is robust to the natural ray spaces, and the generally learned dictionary is still valid to enhance the reconstruction quality of ray space. In the remaining comparison, we only adopt the dictionary learned from "Fuzzy" and "Kuma" ray spaces for performance illustration.

All the recovered EPIs are stacked together to reconstruct the whole ray space. **Figures 3.25** and **3.26** present the generated FVIs from the ray spaces that are reconstructed by using different compressed matrices. The sensing ratio was set as $R = 50\%$. It is shown that the generated FVIs by the learned dictionary have better subjective quality with higher PSNR values than the other two candidates, which is consistent with the result of sparsity comparison in **Figure 3.22**. Furthermore, the reconstruction qualities of the two ray spaces are different; "Kuma" is reconstructed better than "Fuzzy" in the same condition. This is due to the complexity difference between the two ray spaces; "Fuzzy" has more objects with different depths than "Kuma", resulting in more complex structures in the ray space.

Finally, we show a comparison among the three cases with respect to the computational time required for the recovery of one EPI. We used a PC with a 3.20GHz Intel(R) Core(TM) i7 CPU and 3.0 GB main memory, and developed the software using Matlab 2013a without parallelization.



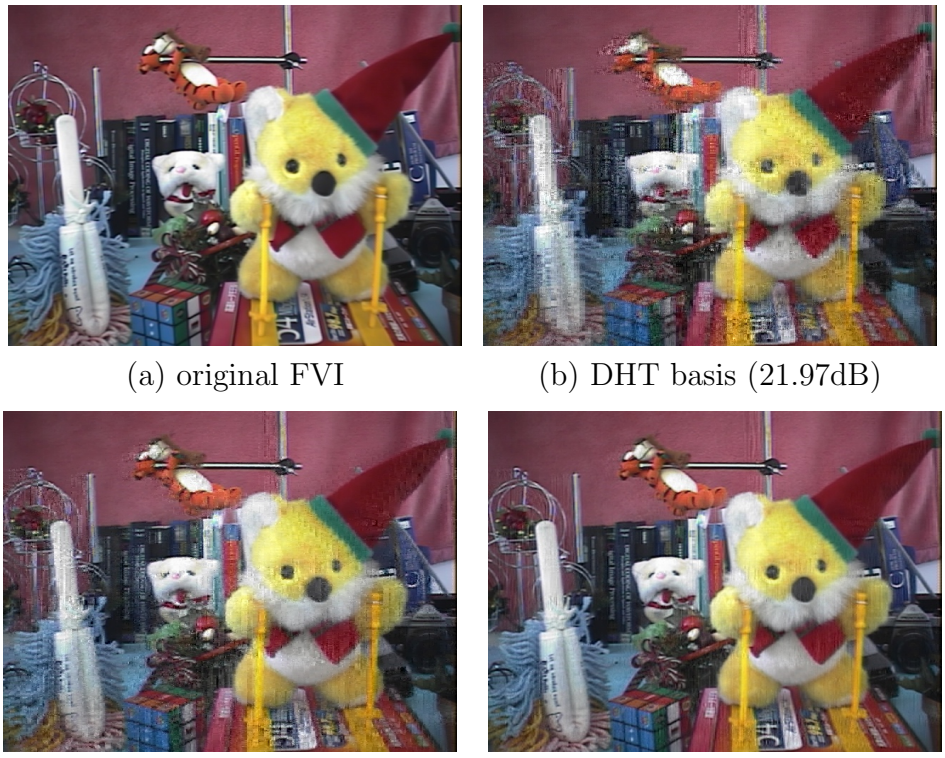
(a) "Fuzzy" ray sapce (b) "Kuma" ray space

Figure 3.24: The performance comparison between two dictionaries learned from different training data sets

As for "Fuzzy" EPI, the orthogonal basis, the designed 2D Gabor dictionary [101], and the learned dictionary consume 1.03s, 69.89s and 12.62s, respectively. As for "Kuma" EPI, the computational time of three candidates are 1.11s, 55.84s, and 12.64s respectively. Obviously, the orthogonal basis requires the smallest computational cost but it produces the worst reconstruction result. Compared to the 2D Gabor dictionary, the learned dictionary requires less computational cost and achieves better reconstruction quality.

3.5 Summary and Discussion

In this chapter, the compressive acquisition model of ray space was firstly mentioned. Afterwards, in the reconstruction part, based on the assumption that if a signal can be more sparsely represented in certain dictionary it can be better reconstructed from incomplete sampled measurements, we proposed two types of dictionaries. One is generated from tunning parameter values of Gabor function, and the other is generated from dictionary learning method. Both of two dictionaries achieved better performance in terms of sparse representation of EPI and reconstruction of ray space



(a) original FVI (b) DHT basis (21.97dB)

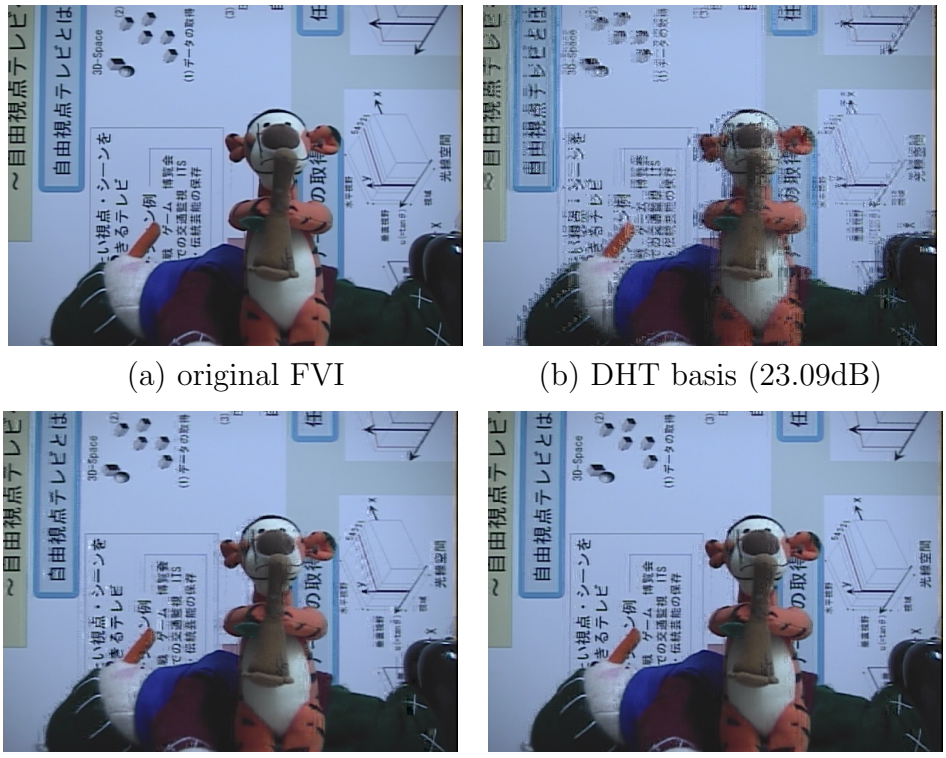
(c) 2D Gabor dictionary (24.39dB) (d) Learned dictionary (30.47dB)

Figure 3.25: "Fuzzy" FVI from reconstructed ray space by using different compressed matrices

than the orthogonal bases. The main reason is that the number of atoms in dictionary is enlarged, and the features of EPI can be well grasped by either 2D Gabor dictionary or learned dictionary.

In comparison between the two dictionaries, the 2D Gabor dictionary is an implicit dictionary and it saves memory in storage because only several parameters are necessary. In contrast, learned dictionary is an explicit dictionary and it requires large memory to store a whole matrix. However, dictionary learning method is much more efficient and the size of dictionary is more compact. Thus, if the reconstruction quality is only taken into consideration, dictionary learning method is a better option.

However, if the reconstruction speed is taken into consideration, over-complete dictionary, such as 2D Gabor and learned dictionary, have no



(a) original FVI (b) DHT basis (23.09dB)

(c) 2D Gabor dictionary(30.03dB) (d) Learned dictionary (34.62dB)

Figure 3.26: "Kuma" FVI from reconstructed ray space by using different compressed matrices

advantage at all compared to orthogonal basis, such as 2D DCT, although the learned dictionary is better than the 2D Gabor dictionary.

For one thing, the size of overcomplete dictionary is larger, thus it results in higher computation in reconstruction. For another thing, orthogonal basis also allows fast computing algorithm, which helps it to win the game of running speed.

Therefore, we would like to adopt orthogonal basis as the compressed matrix for fast running speed, and try to get improvement in reconstruction quality of ray space by modifying other parts in reconstruction procedure. This idea leads us to the next chapter where we will find other priori information of ray space and discuss the question in detail.

Chapter 4

Reconstruction of Ray Space by Statistically Weighted Model

In the previous chapter, we have adopted sparse representation concept to enhance the reconstruction quality of ray space from incomplete sampled measurement. Two designed dictionaries worked quite well because sparser representation could be obtained based on the assumption that sparsity was an important prior in ray space representation and reconstruction. However, considering practical application, the slow speed of reconstruction by learning based method is not suitable and acceptable.

Therefore, in this chapter, we attempt to look the problem in another perspective. In order to achieve better reconstruction quality and faster reconstruction speed, we propose a statistically weighted model in the reconstruction of compressively sampled ray space. This model can explore the structure of ray space data in an orthogonal basis, and integrate this structure into the reconstruction of ray space. In the experiment, the proposed model can achieve much better reconstruction quality for both 2D image patch and 3D image cube cases. Especially in a relatively low sensing ratio, about 10%, the proposed method can still recover most of the low frequency components which are of more significance for representation of ray space data. Besides, the proposed method is almost as good as the state of the art technique, dictionary learning based method, in terms

of reconstruction quality, and the reconstruction speed of our method is much faster. Therefore, the proposed method in this chapter achieves better trade-off between reconstruction quality and reconstruction time, and is more suitable in the practical applications.

4.1 Motivation and Key Point

4.1.1 Motivation

In the application of compressed sensing to the reconstruction of ray space, a random sensing matrix is usually adopted, and recently, several works have started to consider other optimized sensing strategies [96, 102]. In this chapter, however, we still adopt random matrix as the sensing matrix.

Next, there are two options in the selection of compressed matrix. One is an orthogonal basis, such as a DCT basis, and the other is an overcomplete dictionary, which can be carefully designed by tuning a variety of parameters or learned from training data. Thus, a trade-off between effectiveness and efficiency appears. The orthogonal basis-based method enjoys fast computation in reconstruction and the compressed matrix does not rely on any training data. However, the reconstruction quality by this method is not quite satisfying because the structure of basis is fixed and can not be adaptive to specific features of target signal. On the other hand, the overcomplete dictionary is more widely and successfully adopted in the application of compressed sensing to ray space because of the high quality of reconstruction. In last chapter, a structured parametric overcomplete dictionary was designed to be integrated in compressed sensing, and better reconstruction quality was achieved than the orthogonal basis. Furthermore, besides the work in last chapter, in paper [96, 102–104], dictionary learning method was also proposed to replace orthogonal basis and significant improvement of reconstruction quality was reported. The standing point actually was that if one signal could be more sparsely represented in certain dictionary, better reconstruction quality would be expected. However, the performance by using dictionary learning-based method heavily depends on similarity between the learned dictionary and

the data that we want to reconstruct. To make things worse, as far as we know, most works above have sacrificed the reconstruction speed to get better reconstruction quality in the trade-off problem. However, slow reconstruction speed has become a great limitation in further practical application, especially in portable devices.

Therefore, the proposed method in this chapter is still based on an orthogonal basis (2D-DCT or 3D-DCT) for its fast speed in reconstruction, and we attempt to enhance the reconstruction quality by considering other prior information. In fact, it has to be noted that all the works mentioned above (dictionary learning-based method) payed no attention to explore the prior probability of each non-zero element in final sparse solution. In other words, each atom in the learned dictionary was deemed to have the same probability to be selected in the reconstruction. One reasonable explanation is that there is no structure in one learned dictionary and the positions of non-zero elements in sparse representation by overcomplete dictionary almost show uniform distribution, thus the probability of atoms in dictionary has almost no difference. However, the fact is changed as an orthogonal basis (especially DCT basis) is adopted. Since ray space data are natural signal, the non-zero coefficients present clear amplitude structure in frequency domain where the amplitude of low frequency components are larger and the amplitude of high frequency components are smaller. This amplitude structure can be used in the adjustment of prior information of each non-zero element in sparse solution, and better reconstruction quality can be expected.

4.1.2 Key Point

In order to employ this prior information, a modification in reconstruction process is necessary. Usually, as mentioned in chapter 2, in the reconstruction of compressed sensing, most works resort to l_0 norm optimization which could be approximated by Orthogonal Matching Pursuit (OMP) [61] or plain-l_1 norm optimization which is also referred as Basis Pursuit (BP) [105]. Recently, the author in paper [106] has also proposed to adopt a weighted matrix based on iterative method in the reconstruc-

tion, named as re-weighted method. However, the core idea in that work was to enhance the sparsity of solution, but other prior information still remained untouched. Besides, this method is not suitable to high dimensional data and the weight that is iteratively generated is heavily relied on the accuracy of sparse solution. In addition, the reconstruction will be quite time-consuming as iteration is conducted many times. The proposed method is similar as the work in paper [107], but obvious distinctions can be found in three major places at least. In the first place, the work in [107] was dedicated to the problem of compressed video, and the probability model was established by previous frame. In contrast, our model is blind to target ray space, and is abstracted from the statistical property of training data that are totally different from the target data. The second difference is that in the work [107] the probability was based on the position of non-zero elements in sparse solution while the probability in our work has strong relation with the amplitude of coefficients in DCT domain. The last but not the least, in the work [107] the prior assumption was that the position change of non-zero element in sparse solution was quite smooth whereas in our proposal we believe the importance (energy concentration) of different non-zero elements in frequency domain is quite varied and has certain structure.

Therefore, the key points in this chapter are basically two manifolds. First of all, we propose a statistically weighted reconstruction model which modifies the prior probability of non-zero elements in sparse solution. Thus, we can enhance the reconstruction quality, even better than dictionary learning-based method, while the reconstruction time is greatly reduced. In the second place, different from other methods of designing weighted matrix, the proposed weighted matrix has strong relation with amplitude of sparse coefficient and we also present a solution to solve the weighted l_1 norm reconstruction.

4.2 Statistically Weighted Model Design

Before we present the proposed method, the reconstruction of ray space is written as

$$\hat{x} = \arg\min_{x} \| z_{P\times 1} - \Psi_{P\times N}\Phi_{N\times N}x_{N\times 1} \|_{l_2}^2 + \lambda \| x_{N\times 1} \|_{l_1}, \tag{4.1}$$

where the sparse representation $\hat{x}_{N\times 1}$ is firstly estimated from the measurements $z_{P\times 1}$. Later, the signal $\hat{y}_{N\times 1}$ is recovered by $\hat{y}_{N\times 1} = \Phi_{N\times N}\hat{x}_{N\times 1}$. Please note that N is the dimension of processing unit of ray space and P is the number of measurements for this processing unit. Normally, $P < N$ so that the acquisition is compressive one.

4.2.1 Weighted Matrix Design

In this subsection, the proposed model is firstly targeted to 2D image patch, and further we extend the scope to 3D image cube.

The proposed model is a modification of Eq. (4.1), and the target signal $y_{N\times 1}$ corresponds to each non-overlapped 2D image patch of a ray space. The key point of our model is the adjustment of prior probability of each non-zero element in sparse solution. Normally, it is observed that the amplitude of sparse representation of ray space in DCT domain shows certain structure, where the high frequency components are always getting smaller value while the low frequency components tend to have larger amplitudes. However, in the conventional reconstruction method, such as plain l_1 optimization, it only minimized the sum of absolute value of each non-zero element in sparse solution. Thus, high frequency components are more likely to be kept as the non-zero elements in the sparse solution in reconstruction process, and this is totally contradictory with the fact that the low frequency components of an image are much more significant than the high frequency components in sparse representation. Therefore, the modification on the prior probability model is necessary.

We propose to design the corresponding weights to change the prior probability for different frequency coefficients to encourage the low frequency components and to constrain the high frequency components in optimization process. In the specific design of the weighted matrix, a

statistics is conducted from training data, and hence we name the proposed model as statistically weighted model.

First of all, several $\sqrt{N} \times \sqrt{N}$ non-overlapped 2D image patches from ray spaces are firstly prepared as a set of training data. Each 2D image patch, $y^{2D}_{\sqrt{N}\times\sqrt{N}}$, is unfolded to be one column vector $y_{N\times 1}$ in the column-major scan order. These vectors are used to form the training signal matrix $Y_{N\times L} = [y_1, y_2, \ldots, y_L]$, where L is the number of columns in $Y_{N\times L}$ and $y_i \in R^N$. Similarly, the coefficients given by $x_{N\times 1} = \Phi^{-1}_{N\times N} y_{N\times 1}$ are gathered to form $X_{N\times L} = [x_1, x_2, \ldots, x_L]$. Please note that $\Phi^{-1}_{N\times N}$ corresponds to 2D-DCT and we only conduct statistical operation on the coefficients of training data in 2D-DCT domain, and this is totally different from any other training methods.

Here, we illustrate the derivation of the matrix $\Phi^{-1}_{N\times N}$ and $\Phi_{N\times N}$ that correspond to forward (analysis) and backward (synthesis) 2D-DCT for the original image patch. The 2D-DCT of a 2D image patch, $y^{2D}_{\sqrt{N}\times\sqrt{N}}$, is written as

$$x^{2D}_{\sqrt{N}\times\sqrt{N}} = F_{\sqrt{N}\times\sqrt{N}}\, y^{2D}_{\sqrt{N}\times\sqrt{N}}\, (F_{\sqrt{N}\times\sqrt{N}})^T, \tag{4.2}$$

where $x^{2D}_{\sqrt{N}\times\sqrt{N}}$ includes the coefficients in 2D-DCT domain, $F_{\sqrt{N}\times\sqrt{N}}$ represents 1D-DCT in the vertical direction, and $(F_{\sqrt{N}\times\sqrt{N}})^T$ represents 1D-DCT in the horizontal direction. As illustrated in **Figure 4.1**, the equivalent representation is given by $x_{N\times 1} = \Phi^{-1}_{N\times N} y_{N\times 1}$, where $x_{N\times 1}$ corresponds to $x^{2D}_{\sqrt{N}\times\sqrt{N}}$, and $y_{N\times 1}$ corresponds to $y^{2D}_{\sqrt{N}\times\sqrt{N}}$. In this case, $\Phi^{-1}_{N\times N}$ is generated by two 1D-DCT matrices by Kronecker product, written as

$$\Phi^{-1}_{N\times N} = F_{\sqrt{N}\times\sqrt{N}} \otimes F_{\sqrt{N}\times\sqrt{N}}, \tag{4.3}$$

where "$\otimes$" is the Kronecker product. Similarly, $\Phi_{N\times N}$ is given by

$$\Phi_{N\times N} = (F_{\sqrt{N}\times\sqrt{N}})^T \otimes (F_{\sqrt{N}\times\sqrt{N}})^T. \tag{4.4}$$

Then, we take the averaged spectrum energy of each row of $X_{N\times L}$ as the indication in the design of weighted matrix because each row of $X_{N\times L}$ corresponds to the same frequency component in 2D-DCT domain for different sample patches, representing the statistical property of sparse

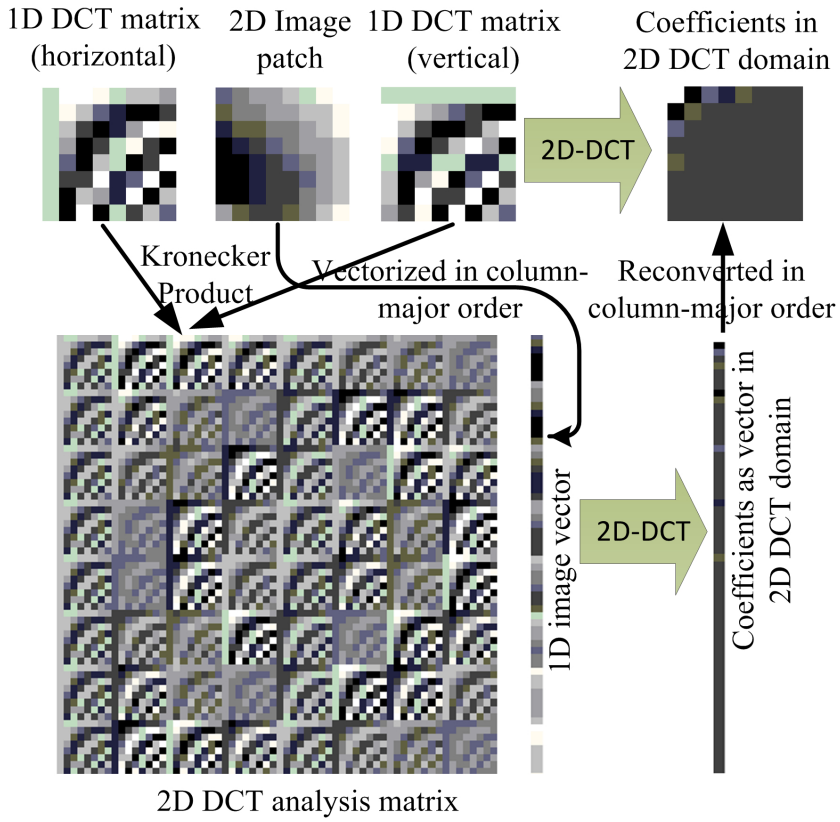
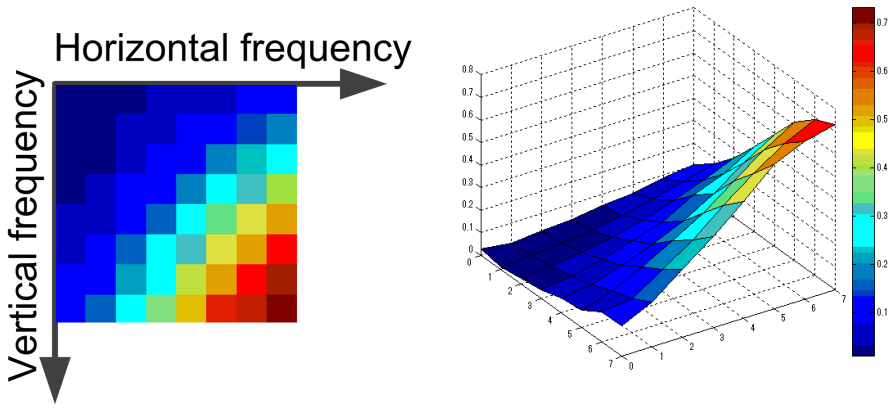


Figure 4.1: The illustration of equivalence between two types of 2D-DCT

representation. Thus, we can obtain the statistically averaged value of spectrum vector $S^T = [s_1, s_2, ...s_N]^T$, where

$$s_i = (\frac{1}{L}\sum_{j=1}^{L} x_{i,j}^2)^{\frac{1}{2}}, i \in [1, N]. \tag{4.5}$$

Next, the weighted matrix W can be designed by adopting the inverse value of the corresponding average spectrum, written as $W_{t,t} = \frac{1}{s_t+\epsilon}, t \in [1, N]$, where ϵ is a very tiny constant to ensure the denominator not to be zero. One example of the weighted matrix for the case of 2D image patch is visualized in **Figure 4.2**. The weighted matrix W is reconverted to original 2D frequency domain (horizontal and vertical) of image patch in the order of column-major scan, where $W_{1,1}$ is placed in the upper-left corner while $W_{64,64}$ is placed in the bottom-right corner. Finally, the



(a) Visualization (b) Surface

Figure 4.2: Visualization and surface of the statistically-weighted matrix for l_1 norm optimization.(2D image patch, $N = 64$)

reconstruction by adopting the statistical weighted model can be written as

$$\hat{x} = arg\min_{x} \| z_{P\times 1} - \Psi_{P\times N}\Phi_{N\times N}x_{N\times 1} \|_{l_2}^2 + \lambda \| W_{N\times N}x_{N\times 1} \|_{l_1}, \quad (4.6)$$

4.2.2 Solution of New Model

In the following, we present how to solve the optimization by the new model. Since the elements along the diagonal of designed W are non-zero, there exists the inverse of W, written as W^{-1}. Next, we introduce x' that satisfies $x = W^{-1}x'$, and the solution estimation is equivalently written as

$$\hat{x} = \arg\min_{x'} \| z_{P\times 1} - D'_{P\times N}x'_{N\times 1} \|_{l_2}^2 + \lambda \| x'_{N\times 1} \|_{l_1}, \quad (4.7)$$

where $D'_{P\times N} = \Psi_{P\times N}\Phi_{N\times N}W^{-1}_{N\times N}$. Note that Eq. (4.7) is equivalent to Eq. (4.1). Therefore, both of the optimizations can be solved by same tool. The only difference is that $D'_{P\times N}$ is used in the optimization problem and the final result $x_{N\times 1}$ is obtained by compensation of $W^{-1}_{N\times N}$ to the solution $x'_{N\times 1}$. The whole procedure is illustrated in **Procedure-1**.

Based on the procedure mentioned above, we extend the target signal to 3D cube of ray space. Since a ray space includes three dimensional data

Procedure-1: The whole simulation procedure of compressively sampling and reconstruction of ray space

Compressively sampling stage:

Input: A non-overlapped image patch from ray space, $y_{N\times 1}$, Sensing matrix $\Psi_{P\times N}$.

Output: The measurement $z_{P\times 1}$ is directly obtained by $z_{P\times 1} = \Psi_{P\times N} \times y_{N\times 1}$.

Reconstruction stage:

1: Weighted matrix design

1) Training data selection: $Y_{N\times L}^{training}$ from Stanford light field archives,

2) Analysis compressed matrix generation: $\Phi_{N\times N}^{-1} = F_{\sqrt{N}\times\sqrt{N}} \otimes F_{\sqrt{N}\times\sqrt{N}}$,

3) Forward transform: $X_{N\times L} = \Phi_{N\times N}^{-1} Y_{N\times L}^{training}$,

4) Statistical operation on obtained coefficients: $S^T = [s_1, s_2, ...s_N]^T$, where $s_i = (\frac{1}{L}\sum_{j=1}^{L} x_{i,j}^2)^{\frac{1}{2}}, i \in [1, N]$,

5) Weighted matrix generation: $W_{t,t} = \frac{1}{s_t+\epsilon}, t \in [1, N]$. If s_t is non-zero, ϵ can be ignored.

2: Solution of optimization

6) x' is introduced to satisfy $x = W^{-1}x'$. Thus, the problem is equivalently presented as $\hat{x'} = arg\min_{x'} \parallel z_{P\times 1} - \Psi_{P\times N}\Phi_{N\times N}W_{N\times N}^{-1}x'_{N\times 1} \parallel_{l_2}^2 + \lambda \parallel x'_{N\times 1} \parallel_{l_1}$

7) Final result: $\hat{x} = W^{-1}\hat{x'}$.

at least[1], if the redundancy in the third dimension is considered, the reconstruction quality will be enhanced. Similar as the 2D case, 3D extension involves the vectorization of 3D cube and 3D-DCT matrix generation.

Non-overlapped cube with size $\sqrt[3]{N} \times \sqrt[3]{N} \times \sqrt[3]{N}$ is taken as the processing unit. The total number of pixel in one 3D cube is N, and the 3D-DCT matrix is generated from three 1D-DCT matrices by Kronecker product, written as

$$\Phi_{N\times N}^{-1} = (F_{\sqrt[3]{N}\times \sqrt[3]{N}})^T \otimes (F_{\sqrt[3]{N}\times \sqrt[3]{N}})^T \otimes (F_{\sqrt[3]{N}\times \sqrt[3]{N}})^T. \tag{4.8}$$

As the compressed matrix is generated, each cube is unfolded as a vector for the following procedure, which is the same as the 2D image patch case.

4.3 Experimental Results

4.3.1 Implementation of Sensing Process

Before presenting the experimental results, we demostrate the implementation in the sensing part. Since the DC component contains most of the energy of ray space, it is captured individually in the sensing stage, and the remaining AC components are multiplexed as measurements by sensing matrix Ψ. Similar as previous subsection, we present the sensing strategy of 2D image patch and 3D image cube respectively.

In the sensing process of a 2D image patch, a matrix C^{2D} is designed to take the remaining AC parts out of the target signal. The matrix C^{2D} is represented as $C^{2D}_{N\times N} = I_{N\times N} - \frac{1}{N}E_{N\times N}$, where $I_{N\times N}$ is an identity matrix and $E_{N\times N}$ is a matrix whose elements are all one. Thus, the whole sensing procedure is formulated as

$$\begin{bmatrix} z^{dc}_{1\times 1} \\ z_{P\times 1} \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\cdot \mathbf{1}^T_{1\times N} \\ \tilde{\Psi}_{P\times N} C^{2D}_{N\times N} \end{bmatrix} \cdot y_{N\times 1}, \tag{4.9}$$

where $\Psi_{P\times N}$ is a randomly generated matrix, and $\mathbf{1}_{1\times N}$ means one vector whose elements are all one. The complete sensing ratio is defined as $\tilde{r} =$

[1]We only take horizontal direction in consideration of angular resolution. Of course, our work can be easily extended to process a 4D ray space with both horizontal and vertical angular resolutions.

$(P+1)/N$. However, as N is large, r is almost same as $\tilde{r}$. Therefore, in the following experiments, we only take the AC components, $z_{P\times 1}$, into consideration, and the sensing ratio $r = P/N,\ 0 < r < 1$ is adopted.

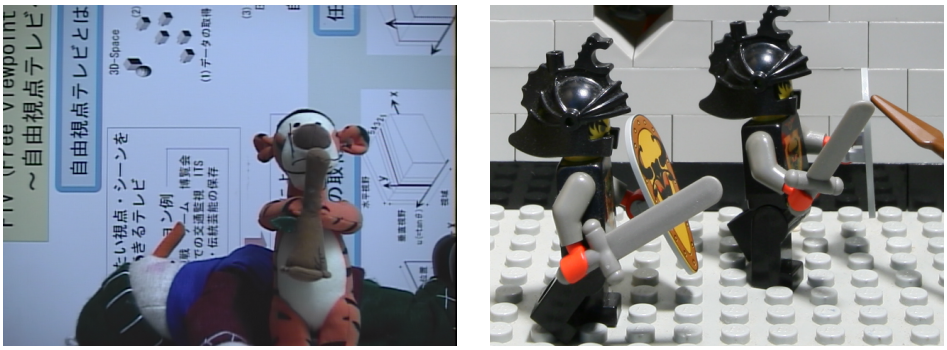
In the sensing process of a 3D cube of ray space, similar sensing implementation is conducted. The only difference is that for 3D cube case, we observe M DC components, where M is the number of angular images in this cube. Kronecker product is adopted in the construction of sensing matrix, and one matrix C^{3D} is also defined to distinct the AC and DC components, where $C^{3D}_{(NM)\times(NM)} = C^{2D}_{N\times N} \otimes I_{M\times M}$. Therefore, for the 3D cube case, the sensing procedure is formulated as

$$\begin{bmatrix} z^{dc}_{M\times 1} \\ z_{P\times 1} \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \cdot \mathbf{1}^{\mathbf{T}}_{\mathbf{1}\times\mathbf{N}} \otimes I_{M\times M} \\ \Psi_{P\times(NM)}(C^{2D}_{N\times N} \otimes I_{M\times M}) \end{bmatrix} \cdot y_{(NM)\times 1}. \tag{4.10}$$

Thus, similar as 2D case, the complete definition of the sensing ratio is $\tilde{r} = (P+M)/(NM)$. However, we roughly adopt $r = P/(NM),\ 0 < r < 1$ as the sensing ratio for in the following experiments.

4.3.2 Setups for Experiments and Parameter Selection

In this subsection, we present our experimental setups and parameter selection. We use a PC with a 3.20 GHz Intel(R) Core(TM) i7 CPU and 3.0 GB main memory to develop the software using Matlab 2013a without parallelization, where l_1 Magic package [70] is adopted for obtaining sparse solution in reconstruction. In the test, we take our own data base "Kuma" ray space and another sequence "Knight" from the Stanford archive with the resolution of $640\times480\times8$. Both of the ray spaces are shown in **Figure 4.3**. In the acquisition, we adopt a random sensing matrix $\Psi_{P\times N}$, whose entries are extracted from the Gaussian distribution, $\psi_i \in N(0,\sigma^2)$, to multiplex the AC components of ray space. Considering the training data set for the weighted matrix generation, five sequences from the Stanford archives [108] are selected, and each sequence includes 8 angular images with the resolution of 640×480 from different perspectives. In the 2D case, we select 5000 random patches from each sequence to form the training



(a) "Kuma" ray space (b) "Knight" ray space

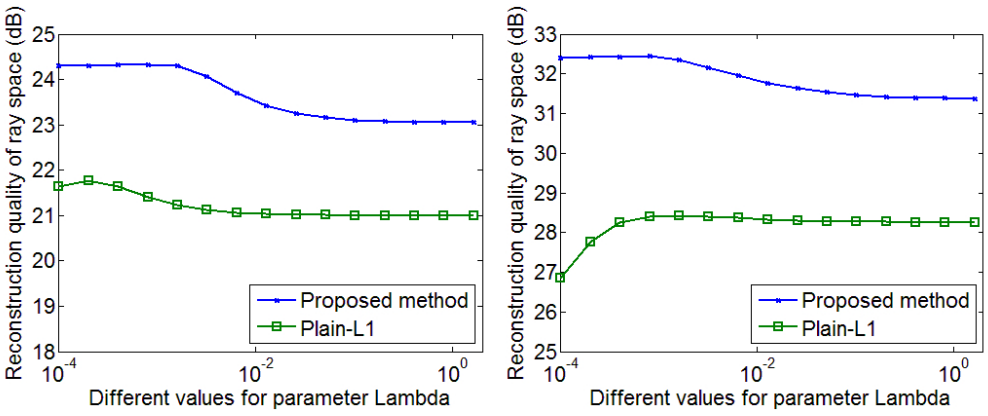
Figure 4.3: One perspective image for two testing ray spaces

data set. In the 3D case, we select 600 random cubes from each sequence to form the training data set.

In the reconstruction, the parameter λ is set as 0.005, and the processing unit is selected as a 8×8 patch and a $8 \times 8 \times 8$ cube for the 2D and the 3D case, respectively. Next, we present the details for the selection of parameter λ and patch size. In order to save page, we take 2D image case as example.

In the first place, we discuss the effect of parameter λ, which plays the role to balance two terms in Eq. 4.6. Two sensing ratios, $r = 0.1$ and $r = 0.5$, were selected for "Knight" ray space. The processing unit was set to 8×8 patches. We present the variety of reconstruction quality as the value of λ is changed in **Figure 4.4** ($\lambda = 10^{-4} \times 2^i$, where $i \in [0, 15]$). It is found that the proposed method consistently outperforms the plain l_1 optimization for a wide range of λ. Moreover, the reconstruction quality is not so sensitive to the change of λ. The value $\lambda = 0.005$ is a good choice for the proposed and plain l_1 optimization methods. Therefore, we took this value throughout the experiment in Section 4 for simplicity, although the optimal λ might be slightly changed depending on dataset, the weighted matrix W, and the size of processing units.

In the second place, we discuss the selection of the size of processing units. The choices of processing unit size are unlimited, because the unit can take any size. Therefore, we picked up several candidate sizes to find

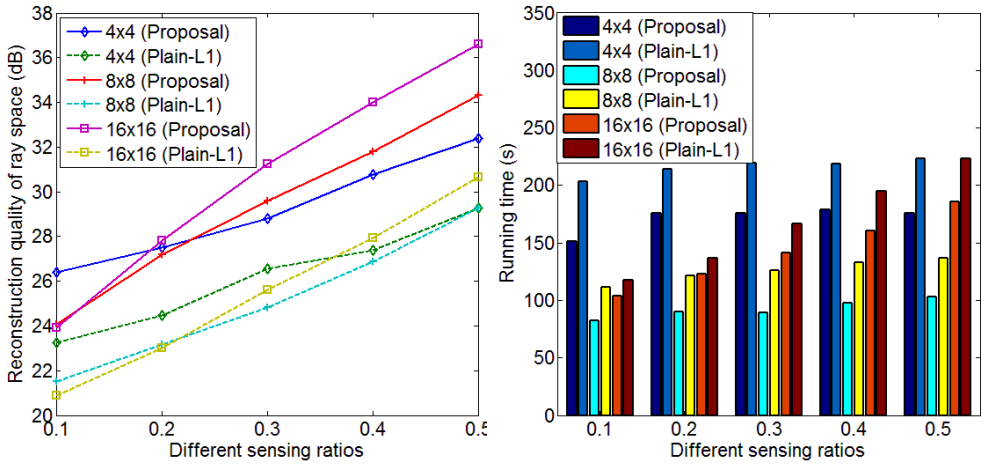


(a) $r = 0.1$ (b) $r = 0.5$

Figure 4.4: The illustration of variety of reconstruction quality as λ is changed for 2D image case ("Knight" ray space)

some tendencies. We used 2D patches whose sizes are 4×4, 8×8, and 16×16. The weighted matrices were computed from the dataset according to the three different patch sizes. We took "Kuma" ray space, and set λ to 0.005 and the sensing ratio r from 0.1 to 0.5. The reconstruction quality is illustrated in **Figure 4.5(a)**. Clearly, reconstruction quality of the proposed method is consistently better than the plain l_1 norm optimization. It is also observed that the smaller size (4×4) performs better at low sensing ratio while the larger size (16×16) performs better at high sensing ratio. The performance of 8×8 patch size is marginal between them for the proposed method. The computational time is illustrated in **Figure 4.5(b)**. If the patch size is small, the total number of patches increases. While the patch size is large, the computational cost for each patch increases. It is observed from the graph that 8×8 achieved the best balance between the number of patches and computational cost for each patch, resulting in minimum computational cost in total. Therefore, 8×8 is a reasonable choice in terms of reconstruction quality and computational time. For the 3D cube case, we simply adopted $8 \times 8 \times 8$ as an example although there are numerous possibilities.

In the next subsection, we present the numerical results to illustrate the effectiveness and efficiency of the proposed model.

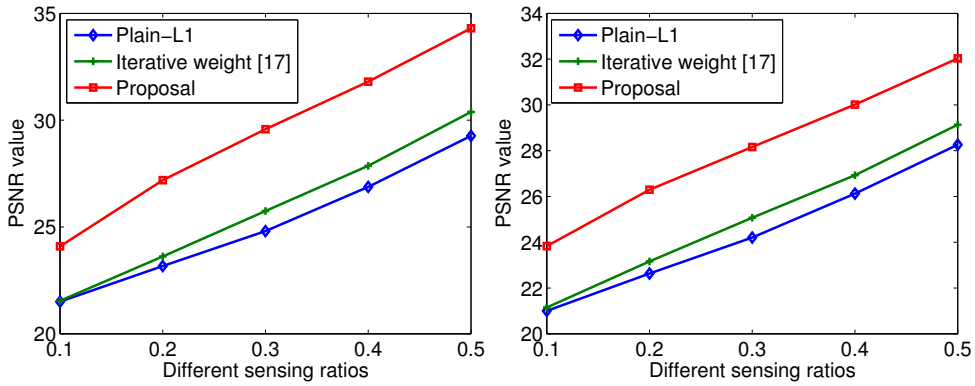


(a) Reconstruction quality (b) Computation time

Figure 4.5: The illustration of reconstruction quality and computation time at different patch size for 2D image case ("Kuma" ray space)

4.3.3 Numerical Results

In the first place, the comparison between our proposal (Weighted-l_1), plain-l_1 and the method from [106] (Iterative weight) are presented. (Please note that as the implementation from [106], only two iterations (computation for sparse solution) are conducted, where the first one is used to generate the weighted matrix and the second one is used to get the final sparse solution.) As this is a simulation work, we have obtained the ray spaces as original data, and compressively sampling operation is conducted with different sensing ratios from 0.1 to 0.5 ($0.1 \leq r \leq 0.5$). We take Peak of Signal Noise Ratio (PSNR) value to evaluate the distortion of each reconstruction method. The averaged PSNR value of the recovered ray spaces at different sensing ratios are presented in **Figure 4.6**. As for the method in [106], the weight is heavily relied on the accuracy of sparse solution. Thus, as sensing ratio is relatively low, the accuracy of initial sparse solution is also low, and the improvement by the addition of weight is quite limited. As sensing ratio increases, however, especially over the ratio $r = 0.3$, the accuracy of the first sparse solution becomes higher, resulting in better performance than plain l_1 norm optimization. For our

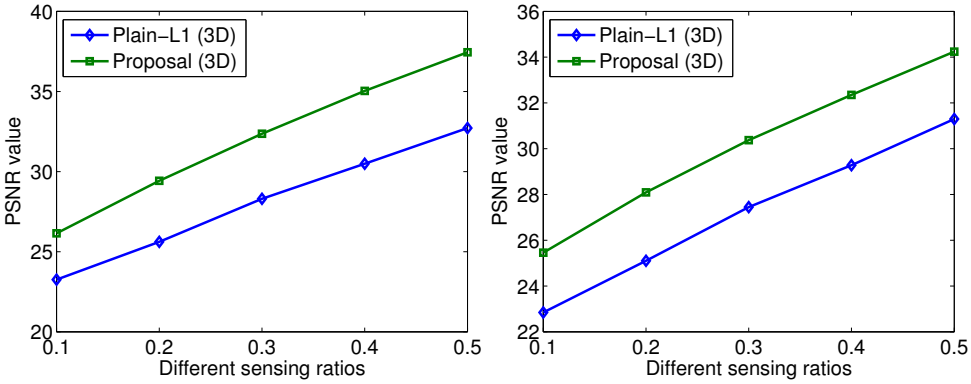


(a) "Kuma" ray space (b) "Knight" ray space

Figure 4.6: The comparison of reconstruction performance at different sensing ratios for two ray spaces (2D case)

proposed method, it always has great enhancement in reconstruction quality for both testing ray spaces, compared to the other two candidates. The main reason is that our proposal resorts to the statistical operation and the most significant part, low frequency components, for signal presenting can be recovered even in low sensing ratio. Afterwards, we present the results for the 3D case in **Figure 4.7**. Since the method in [106] is not suitable to high dimensional data, the comparison between our proposal and plain l_1 optimization is illustrated, and obviously our proposal has great improvement. In addition, it is also illustrated that the 3D cube case is much better than the 2D patch case, because 3D cube case exploits three dimensional redundancy.

Next, we compare the reconstruction time of different methods and different types of processing unit. Similar as the previous comparison, the sensing ratio is also selected from 0.1 to 0.5 with the increment of 0.1. Please note that the running time is only regarded as the time to estimate the sparse solution, because this part is the main computation in the whole reconstruction process. The result of 2D case is presented in **Figure 4.8**. Our proposed method can reduce the reconstruction time by about 25%. Besides, for the method in [106], another computation for sparse solution is required, and hence the reconstruction time is higher than the one by the



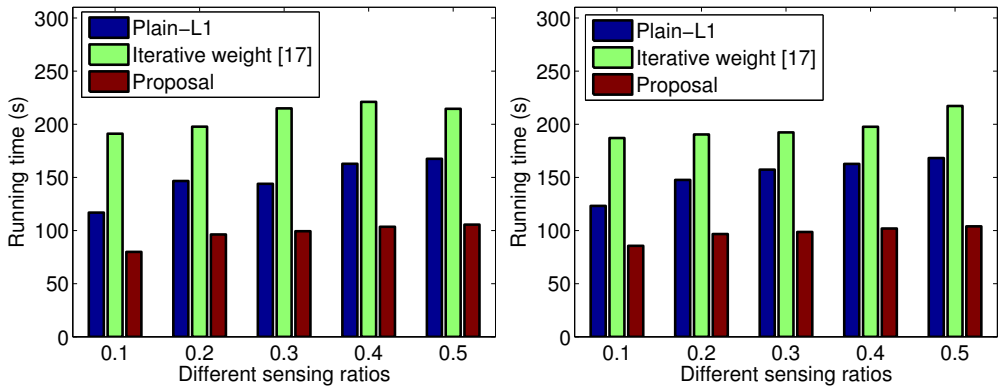
(a) "Kuma" ray space (b) "Knight" ray space

Figure 4.7: The comparison of reconstruction performance at different sensing ratios for two ray spaces (3D case)

plain l_1 norm optimization. As the iteration is extended, the computation time will also increase correspondingly. For 3D cube case illustrated in **Figure 4.9**, the proposed method can still cut the running time by 10% compared to the plain l_1 method. The main reason for the reduction of computational time comes that the addition of weighted matrix promotes faster convergent speed in the optimization.

In the following part, we present the subjective evaluation of the reconstruction quality for both 2D and 3D case. For "Kuma" and "Knight" ray spaces, they are compressively sampled at low sensing ratio 0.1 and high sensing ratio 0.5, respectively. The subjective evaluations are shown in **Figure 4.10, 4.11**. The second row of each figure is the close-ups for comparison in detail. It is observed that the proposed method has improved a lot compared to the conventional plain-l_1 norm optimization method in both 2D and 3D cases. Especially in the case of low sensing ratio, the proposed method can still have an acceptable reconstruction quality, because our method promotes the low frequency component that is of great importance for ray space data.

Finally, the comparison between our proposed method and dictionary learning-based method in terms of reconstruction quality and reconstruction time is presented, respectively. Here, we only present the result of



(a) "Kuma" ray space (b) "Knight" ray space

Figure 4.8: The reconstruction time at different sensing ratios for two ray spaces (2D case)

3D case. For dictionary learning-based method, the set of training data is selected as the same as the one for weighted matrix generation, and, for simplicity, Method of Optimal Direction (MOD) [98] is adopted as learning method with 50 iterations. The size of dictionary is chosen as 64×128 with the redundancy of 2. The comparison of reconstruction quality is illustrated in **Figure 4.12**. For "Kuma" ray space, dictionary learning method has advantage in reconstruction quality while for the "Knight" case, our proposal is better than dictionary learning method. The main reason is that dictionary learning-based method relies heavily on the similarity between training and testing sets. While our proposed method is based on statistical property of training set, thus it is more stable in reconstruction of different ray spaces. Next, the reconstruction time is also presented in **Figure 4.13**, and we can clearly see that the proposed method has great advantage over the dictionary learning-based method.

4.4 Summary and Discussion

In this chapter, a statistically weighted model was proposed to enhance the reconstruction performance of compressively sampled ray space. The basic idea was that the sparsity of signal only presented part of prior

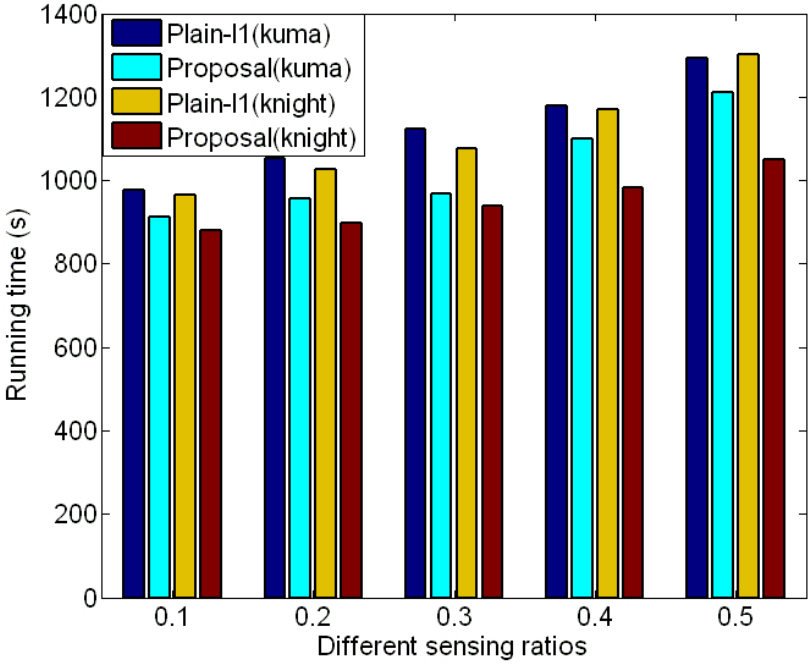


Figure 4.9: The reconstruction time at different sensing ratios for two ray spaces (3D case)

information of signal. In addition, the amplitude structure of non-zero elements in DCT domain was also utilized in our method. A statistical operation was conducted to design a weighted matrix that reflected the amplitude structure, and the weighted matrix was further integrated in the reconstruction process. We discussed 2D patch and 3D cube cases, and the proposed method achieved better results than the conventional one for both cases. In addition, the reconstruction time by proposed method was also reduced, compared to the conventional methods.

Furthermore, another comparison between the proposed method in this chapter and the dictionary learning-based method in the previous chapter was also presented. Clearly, there was a trade-off in aspects of reconstruction quality and reconstruction time between the two methods. For reconstruction quality, dictionary learning-based method still had advantage because the dictionary was adaptively learned. However,

for reconstruction speed, the statistically weighted model was much faster. Therefore, if fast reconstruction time is required in practical application, the statistically weighted model is a better choice.

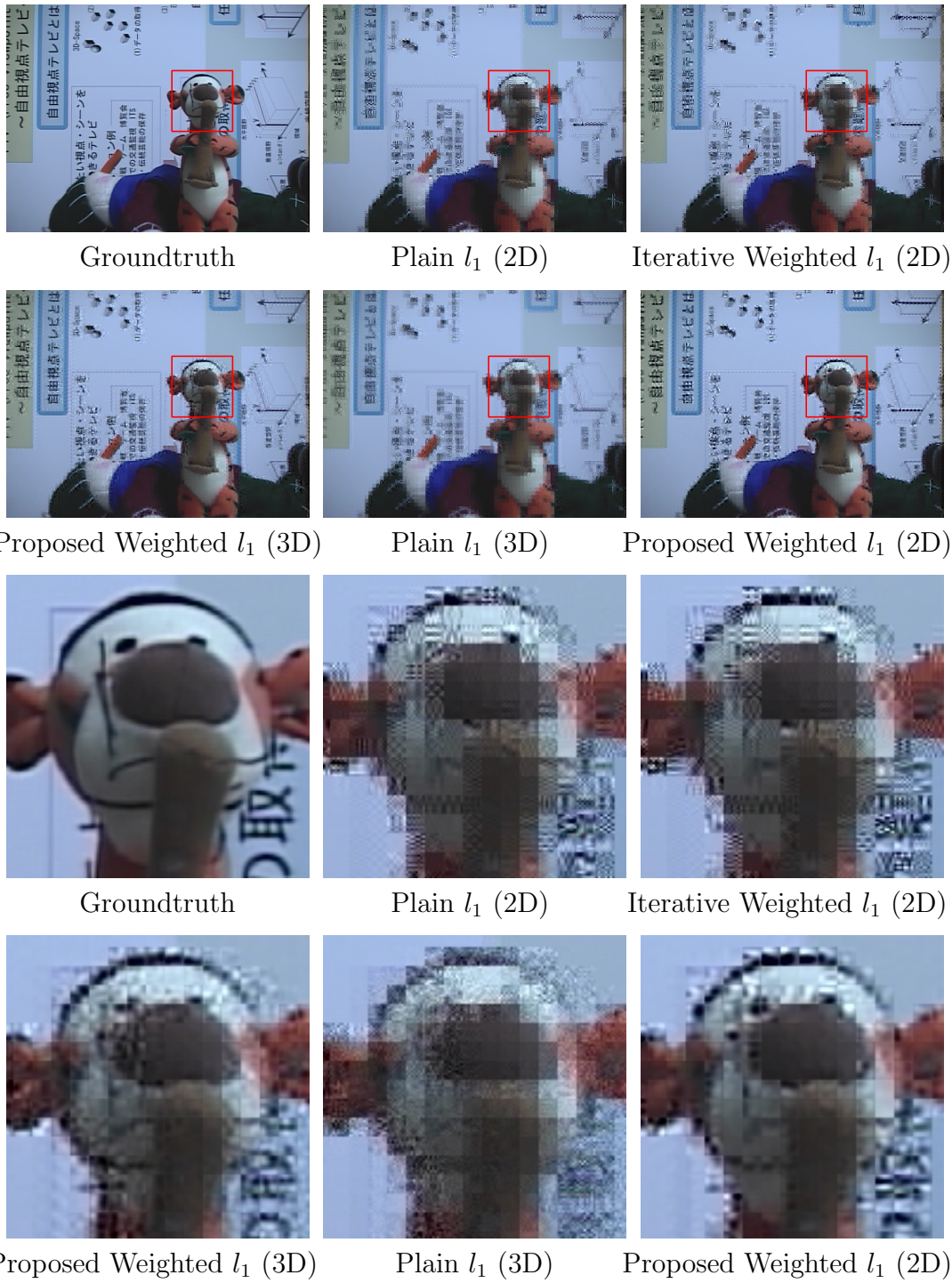
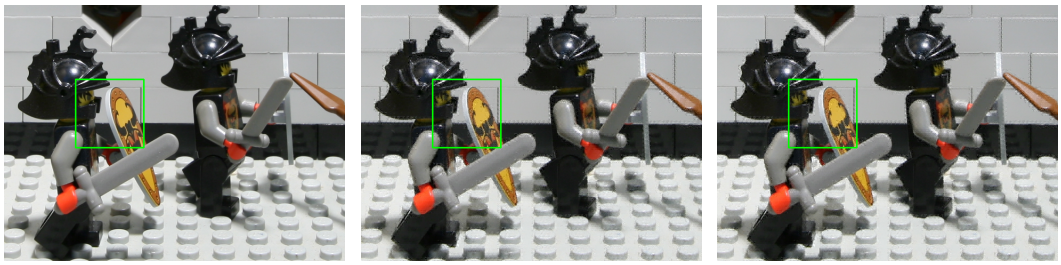


Figure 4.10: Comparison of reconstruction quality among different methods (r=0.1).



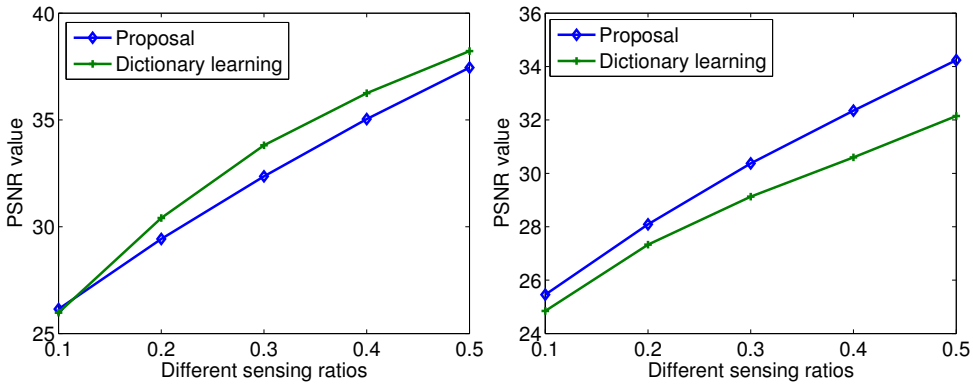
Groundtruth | Plain l_1 (2D) | Iterative Weighted l_1 (2D)

Proposed Weighted l_1 (3D) | Plain l_1 (3D) | Proposed Weighted l_1 (2D)

Groundtruth | Plain l_1 (2D) | Iterative Weighted l_1 (2D)

Proposed Weigthed l_1 (3D) | Plain l_1 (3D) | Proposed Weighted l_1 (2D)

Figure 4.11: Comparison of reconstruction quality among different methods (r=0.5).



(a) "Kuma" ray space (b) "Knight" ray space

Figure 4.12: Comparison of reconstruction quality between proposal and dictionary learning-based method

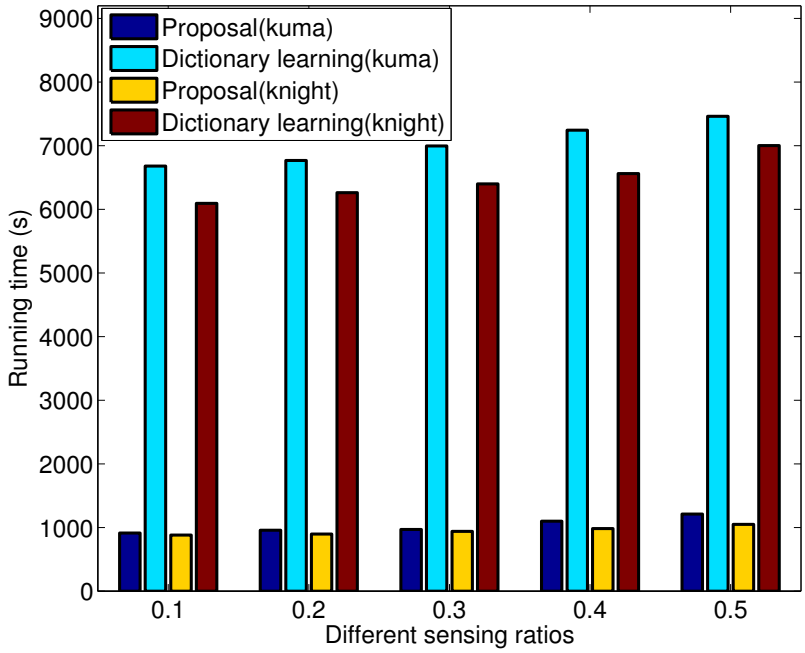


Figure 4.13: Comparison of reconstruction time between proposal and dictionary learning-based method

Chapter 5

Conclusions

In this thesis, based on the main purpose at the first beginning, we have analyzed the problem of compressive acquisition and reconstruction of ray space motivated by the huge data problem in sensing process. In addition, the quantitative relationship between necessary number of measurements and the final reconstruction quality has been explored. Besides, two novel proposals dedicated to enhancing reconstruction quality at fixed sensing ratios have been discussed in previous chapters.

However, due to the time and capability of the author, there are many limitations of the current proposals. In addition, there are also several interesting areas that are worthy of taking proceeding. Combining the limitations and following directions, we point out the important future works.

5.1 What is All About for Conclusion

At the beginning, we posed the question that how could we capture light (ray space) rather than the projection light. Then, in order to describe ray space, the plenoptic function was also introduced for parameterization of ray space. After that, the conception of computational photography and several typical applications of ray space were discussed. Thus, the efficient and effective acquisition of a ray space was quite necessary. By the conventional acquisition method, the acquired data of ray space were quite huge, especially at the dense sampling rate. As the resolution of

ray space increases, the data problem becomes more serious. In addition, even the huge amount of data could be obtained, compression had to be also conducted before data transmission and storage. Therefore, in the acquisition of ray space, we need to integrate acquisition and compression together for efficiently capturing a ray space. Fortunately, the theory of compressed sensing provided substantial theoretical support for our work, and the theory was also roughly introduced so as to enable readers to better understand the following contents.

If the background of ray space and compressed sensing are regarded as the soil for fertilizing seeds, which are our proposals, the related works are regarded as the tools for cultivation. During the related works, the most significant tool, l_p norm optimization for reconstruction of ray space was introduced. In addition, the available systems of acquisition were also discussed in detail so that our contributions could be presented naturally.

The first contribution is related to the combination of sparse coding and reconstruction of ray space. After establishing acquisition model and obtaining compressively sampled measurements of ray space, we endeavored to reconstruct the ray space. Since compressed sensing states that a signal can be recovered from the incomplete samples if the target signal is sparse by itself or sparse in other domains. Thus, we proposed to reconstruct a ray space from compressively sampled measurements by exploring sparsity of ray space. Two types of dictionaries were proposed. One was carefully designed by using Gabor function, thus the generated atoms in 2D Gabor dictionary matched the features of EPI that were basic processing unit of ray space so that better reconstruction quality was achieved. The other dictionary was adaptively developed by learning from a set of training data, thus the atoms from dictionary learning could be shaped to grasp the features of EPI, and hence the reconstruction quality was further enhanced. However, the two dictionaries mentioned above sacrificed reconstruction speed to gain the reconstruction quality, and the slow reconstruction speed became an obstacle for real time applications.

The second contribution attempted to handle the trade-off between reconstruction quality and reconstruction time. We found that the amplitude structure of coefficients of ray space in DCT domain could provide

other priori information in the reconstruction, and we proposed to adopt a weighted matrix to reflect the structure and to integrate the structure in the reconstruction process. In addition, we provided a solution for the new reconstruction process so that the previous optimization solver could be reused. The simulation results showed that the proposed method achieved better reconstruction quality than the bench mark. In the comparison between weighed model-based method and dictionary learning-based method, both methods achieved similar performance (much better than the bench marks) in terms of reconstruction quality. However, the weighed model-based method had great advantage in aspect of reconstruction speed. Therefore, the second contribution gave a better trade-off between reconstruction speed and reconstruction quality.

5.2 What is Still Missing for Future Works

In this section, we pick up several points as the future works. These points are either extension of the current work in the thesis or provide other directions based on the current results.

5.2.1 Extensions of the Contributions

Since there are two main contributions in this thesis, we discuss the corresponding extensions respectively. In the first place, we proposed to integrate overcomplete dictionary (designed or learned) into the reconstruction. Thus, to develop or to search sparser representation of ray space under other dictionary can be one extension of our proposal. As the new dictionary is found, better reconstruction quality is expected. In the second place, in order to achieve a better trade-off between reconstruction quality and reconstruction speed, we proposed a weighted model to grasp the amplitude structure of ray space in DCT domain. However, this weighed model is an off-line model, which means it is static model that has been fixed in reconstruction. If the model can be obtained dynamically during the reconstruction, better reconstruction quality can be achieved, although computational cost has to be slightly increased.

5.2.2 Optimized Sensing Strategy

In this thesis, only the random sensing matrix was adopted in the simulation of compressive acquisition. The main reason was that random sensing matrix had a high opportunity to have the RIP satisfied. However, random sensing matrix was not the optimized sensing matrix, and it was just a general choice. Several researchers have started to explore an optimal sensing strategy for compressive acquisition. Based on the technique of incoherent dictionary learning, an incoherent sensing matrix can be also learned, and better reconstruction result can be expected. In addition, other model such as Gaussian Mixture Model (GMM) has been considered in sequential sensing in order to obtain current optimal measurement based on the previous acquired measurement. Therefore, the research on optimized strategy is an interesting and open topic for designing the most suitable sensing matrix to compressively capture a ray space.

5.2.3 Parallel Computing by GPU

In recent years, the high performance computation by Graphic Processing Unit (GPU) or General Purpose Graphic Processing Unit (GPGPU) has attracted great attentions in numerous researching fields, because the parallel computing is well supported on GPU platform. Since the reconstruction of ray space is quite time consuming on CPU, GPU will be a great alternative to accelerate the reconstruction speed of ray space. More importantly, our processing unit for the reconstruction is the non-overlapped patch of ray space and these patches can be processed individually. Therefore, the processing can be easily transferred to the GPU platform, and the real-time reconstruction on GPU can be possible.

5.2.4 Optical System Design

We put the optical system design in the final part because it is the final target of this research. “Demo or Die” is one popular sentence in engineering area, thus a prototype for illustrating the theories, algorithms, and methods mentioned in the thesis is of great significance for our research. As

mentioned in Chapter 2, there have been several systems available, such as coded aperture, camera array system, reflection mirror system, and etc, for ray space acquisition. However, a more effective, more efficient, and more stable optical system is always highly desired.

5.3 Summary

By mentioning the main work in this thesis and pointing the limitations of the work and future directions, we come to the conclusion of this thesis. However, the work is far from finished. Instead, the work has just started, and there are more spaces for further improvement, more applications for wider development, and more works for deeper involvement.

Bibliography

[1] "In our time archive: Thomas edison," Dec. 2010.

[2] T. Gustavson and G. E. House, *Camera: A History of Photography from Daguerreotype to Digital*, Sterling Innovation, 2009.

[3] H. Ives, "Parallax stereogram and process of making same," 1903.

[4] G. Lippmann, "La photographie integrale," *Academie des Sciences*, vol. 146, pp. 446–551, 1908.

[5] Wetzstein, *Computational Plenoptic Image Acquisition and Display*, Ph.D. thesis, The University Of British Columbia, 2011.

[6] Edward H Adelson and James R Bergen, "The plenoptic function and the elements of early vision," *Computational models of visual processing*, vol. 1, no. 2, 1991.

[7] L. Mcmillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," *Proc. SIGGRAPH' 95*, pp. 39–46, 1995.

[8] Aljoscha Smolic, Karsten Mueller, Nikolce Stefanoski, Joern Ostermann, Atanas Gotchev, Gözde B Akar, Georgios Triantafyllidis, and Alper Koz, "Coding algorithms for 3dtv a survey," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1606–1621, 2007.

[9] Cha Zhang, "Multiview imaging and 3dtv," *IEEE Signal Processing Magazine*, vol. 1053, no. 5888/07, 2007.

[10] Peter Kauff, Nicole Atzpadin, Christoph Fehn, Marcus Müller, Oliver Schreer, Aljoscha Smolic, and Ralf Tanger, "Depth map creation and

image-based rendering for advanced 3dtv services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 217–234, 2007.

[11] Wijnand IJsselsteijn, Huib de Ridder, Roelof Hamberg, Don Bouwhuis, and Jonathan Freeman, "Perceived depth and the feeling of presence in 3dtv," *Displays*, vol. 18, no. 4, pp. 207–214, 1998.

[12] Toshiaki Fujii and Masayuki Tanimoto, "Free viewpoint tv system based on ray-space representation," in *ITCom 2002: The Convergence of Information Technologies and Communications*, 2002, pp. 175–189.

[13] Masayuki Tanimoto, "Ftv (free viewpoint television) creating ray-based image engineering," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. IEEE, 2005, vol. 2, pp. II–25.

[14] Masayuki Tanimoto, Mehrdad Panahpour Tehrani, Toshiaki Fujii, and Tomohiro Yendo, "Free-viewpoint tv," *Signal Processing Magazine, IEEE*, vol. 28, no. 1, pp. 67–76, 2011.

[15] Masayuki Tanimoto, Mehrdad Panahpour Tehrani, Toshiaki Fujii, and Tomohiro Yendo, "Ftv for 3-d spatial communication," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 905–917, 2012.

[16] Cha Zhang and Tsuhan Chen, "A survey on image-based rendering?representation, sampling and compression," *Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1–28, 2004.

[17] Sing Bing Kang, Yin Li, Xin Tong, and Heung-Yeung Shum, "Image-based rendering," *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, no. 3, pp. 173–258, 2006.

[18] SC Chan, Heung-Yeung Shum, and King-To Ng, "Image-based rendering and synthesis," *Signal Processing Magazine, IEEE*, vol. 24, no. 6, pp. 22–33, 2007.

[19] Heung-Yeung Shum, Shing-Chow Chan, and Sing Bing Kang, *Image-based rendering*, Springer, 2008.

[20] Kazuya Kodama, Hiroshi Mo, and Akira Kubota, "All-in-focus image generation by merging multiple differently focused images in three-dimensional frequency domain," in *Advances in Multimedia Information Processing-PCM 2005*. 2005, pp. 303–314, Springer.

[21] Yuichi Taguchi, Keito Takahashi, and Takashi Naemura, "Real-time all-in-focus video-based rendering using a network camera array," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2008*. IEEE, 2008, pp. 241–244.

[22] Mark Harris, "Focusing on everything," *Spectrum, IEEE*, vol. 49, no. 5, pp. 44–50, 2012.

[23] Todor Georgiev, Zhan Yu, Andrew Lumsdaine, and Sergio Goma, "Lytro camera technology: theory, algorithms, performance analysis," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013.

[24] Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum, "Plenoptic sampling," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 307–318.

[25] David L Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[26] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[27] Richard G Baraniuk, "Compressive sensing," *IEEE signal processing magazine*, vol. 24, no. 4, 2007.

[28] Emmanuel J Candès and Michael B Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[29] Holger Rauhut, Karin Schnass, and Pierre Vandergheynst, "Compressed sensing and redundant dictionaries," *Information Theory, IEEE Transactions on*, vol. 52, no. 5, pp. 2210–2219, 2008.

[30] Emmanuel J Candès, Yonina C Eldar, Deanna Needell, and Paige Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[31] Marc Levoy and Pat Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.

[32] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen, "The lumigraph," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 43–54.

[33] Toshiaki Fujii, "Ray space coding for 3d visual communication," in *Picture Coding Symposium'96*, 1996, vol. 2, pp. 447–451.

[34] Robert C Bolles, H Harlyn Baker, and David H Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.

[35] Changyin Zhou and Shree K Nayar, "Computational cameras: convergence of optics and processing," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3322–3340, 2011.

[36] JinLi Suo, XiangYang Ji, and QiongHai Dai, "An overview of computational photography," *Science China Information Sciences*, vol. 55, no. 6, pp. 1229–1248, 2012.

[37] Ramesh Raskar, *Computational Optical Sensing and Imaging-Computational photography*, Optical Society of America, 2009.

[38] Shree K Nayar, "Computational cameras: Approaches, benefits and limits," Tech. Rep., Department of Computer Science, Columbia University, 2011.

[39] Shoji Tominaga, "Spectral imaging by a multichannel camera," *Journal of Electronic Imaging*, vol. 8, no. 4, pp. 332–341, 1999.

[40] ME Gehm, R John, DJ Brady, RM Willett, and TJ Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics Express*, vol. 15, no. 21, pp. 14013–14027, 2007.

[41] E Caroli, JB Stephen, G Di Cocco, L Natalucci, and A Spizzichino, "Coded aperture imaging in x-and gamma-ray astronomy," *Space Science Reviews*, vol. 45, no. 3-4, pp. 349–403, 1987.

[42] Walter R Cook, John M Grunsfeld, William A Heindl, David M Palmer, Thomas A Prince, Stephen M Schindler, and Edward C Stone, "Coded-aperture imaging of the galactic center region at gamma-ray energies," *Astrophysical Journal Letters*, vol. 372, no. 2, pp. L75–L78, 1991.

[43] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CST*, vol. 2, no. 11, pp. 2005, 2005.

[44] Andrew Lumsdaine and Todor Georgiev, "The focused plenoptic camera," in *Computational Photography (ICCP), 2009 IEEE International Conference on*. IEEE, 2009, pp. 1–8.

[45] Todor G Georgiev, Andrew Lumsdaine, and Sergio Goma, "High dynamic range image capture with plenoptic 2.0 camera," *Signal recovery and synthesis*, p. SWA7P, 2009.

[46] Tom E Bishop, Sara Zanetti, and Paolo Favaro, "Light field superresolution," in *Computational Photography (ICCP), 2009 IEEE International Conference on*. IEEE, 2009, pp. 1–9.

[47] Todor Georgiev, Georgi Chunev, and Andrew Lumsdaine, "Superresolution with the focused plenoptic camera," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011, pp. 78730X–78730X.

[48] Tom E Bishop and Paolo Favaro, "Full-resolution depth map estimation from an aliased plenoptic light field," in *Computer Vision–ACCV 2010*. 2011, pp. 186–200, Springer.

[49] Belen Masia, Gordon Wetzstein, Piotr Didyk, and Diego Gutierrez, "A survey on computational displays: Pushing the boundaries of optics, computation, and perception," *Computers & Graphics*, vol. 37, no. 8, pp. 1012–1038, 2013.

[50] Andrew Maimone, Gordon Wetzstein, Matthew Hirsch, Douglas Lanman, Ramesh Raskar, and Henry Fuchs, "Focus 3d: Compressive accommodation display," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 5, pp. 153, 2013.

[51] Douglas Lanman, Matthew Hirsch, Yunhee Kim, and Ramesh Raskar, "Content-adaptive parallax barriers: optimizing dual-layer 3d displays using low-rank light field factorization," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 5, pp. 163, 2010.

[52] Gordon Wetzstein, Douglas Lanman, Wolfgang Heidrich, and Ramesh Raskar, "Layered 3d: tomographic image synthesis for attenuation-based light field and high dynamic range displays," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, pp. 95, 2011.

[53] Gordon Wetzstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar, "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 4, pp. 80, 2012.

[54] Felix Heide, Gordon Wetzstein, Ramesh Raskar, and Wolfgang Heidrich, "Adaptive image synthesis for compressive displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 132, 2013.

[55] HJ Landau, "Sampling, data transmission, and the nyquist rate," *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1701–1706, 1967.

[56] Emmanuel Candès and Justin Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, pp. 969, 2007.

[57] Emmanuel J Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[58] Joel A Tropp, "Greed is good: Algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, 2004.

[59] Michael Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer, 2010.

[60] Stéphane G Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.

[61] Yagyensh Chandra Pati, Ramin Rezaiifar, and PS Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.

[62] Deanna Needell and Joel A Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[63] Rémi Gribonval and Morten Nielsen, "Approximate weak greedy algorithms," *Advances in Computational Mathematics*, vol. 14, no. 4, pp. 361–378, 2001.

[64] Pascal Vincent and Yoshua Bengio, "Kernel matching pursuit," *Machine Learning*, vol. 48, no. 1-3, pp. 165–187, 2002.

[65] Geoff Davis, Stephane Mallat, and Marco Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[66] David G Luenberger, *Introduction to linear and nonlinear programmin*, vol. 28, Addison-Wesley Reading, MA, 1973.

[67] John A Nelder and Roger Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[68] Yurii Nesterov, Arkadii Nemirovskii, and Yinyu Ye, *Interior-point polynomial algorithms in convex programming*, vol. 13, SIAM, 1994.

[69] Guy Desaulniers, Jacques Desrosiers, and Marius M Solomon, *Column generation*, vol. 5, Springer.

[70] Emmanuel Candès and Justin Romberg, "l1-magic: Recovery of sparse signals via convex programming," *URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf*, vol. 4, 2005.

[71] Michael Grant, Stephen Boyd, and Yinyu Ye, "Cvx: Matlab software for disciplined convex programming," 2008.

[72] David L Donoho, Victoria C Stodden, and Yaakov Tsaig, "Sparselab architecture," 2007.

[73] MAT Figueiredo, RD Nowak, and SJ Wright, "Software: Gpsr (gradient projection for sparse reconstruction)," .

[74] Ewout van den Berg, MP Friedlander, G Hennenfent, F Herrmann, R Saab, and O Yılmaz, "Sparco: A testing framework for sparse reconstruction," Tech. Rep., Tech. Rep. TR-2007-20, Dept. Computer Science, University of British Columbia, Vancouver, 2007.

[75] Xiangyang Zhu, Han Ding, and Michael Y Wang, "Form error evaluation: An iterative reweighted least squares algorithm*," *Journal of manufacturing science and engineering*, vol. 126, no. 3, pp. 535–541, 2004.

[76] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy, "High performance imaging using large camera arrays," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, pp. 765–776, 2005.

[77] Jason C Yang, Matthew Everett, Chris Buehler, and Leonard McMillan, "A real-time distributed light field camera," in *Proceedings of the 13th Eurographics workshop on Rendering*, 2002, pp. 77–86.

[78] Toshiaki Fujii, Kensaku Mori, Kazuya Takeda, Kenji Mase, Masayuki Tanimoto, and Yasuhito Suenaga, "Multipoint measuring system for video and sound-100-camera and microphone system," in *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 2006, pp. 437–440.

[79] Kazuya Kodama and Akira Kubota, "Efficient reconstruction of all-in-focus images through shifted pinholes from multi-focus images for dense light field synthesis and rendering," *Image Processing, IEEE Transactions on*, vol. 22, no. 11, pp. 4407–4421, 2013.

[80] Takashi Sakamoto, Kazuya Kodama, and Takayuki Hamamoto, "A novel scheme for 4-d light-field compression based on 3-d representation by multi-focus images," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2901–2904.

[81] EE Fenimore, "Coded aperture imaging: predicted performance of uniformly redundant arrays," *Applied Optics*, vol. 17, no. 22, pp. 3562–3570, 1978.

[82] Todor Georgiev, Ke Colin Zheng, Brian Curless, David Salesin, Shree Nayar, and Chintan Intwala, "Spatio-angular resolution tradeoffs in integral photography," *Rendering Techniques*, vol. 2006, pp. 263–272, 2006.

[83] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 69, 2007.

[84] Todor Georgiev, Chintan Intwala, Sevkit Babakan, and Andrew Lumsdaine, "Unified frequency domain analysis of lightfield cameras," in *Computer Vision–ECCV 2008*. 2008, pp. 224–237, Springer.

[85] Gordon Wetzstein, Ivo Ihrke, and Wolfgang Heidrich, "On plenoptic multiplexing and reconstruction," *International journal of computer vision*, vol. 101, no. 2, pp. 384–400, 2013.

[86] Anat Levin, Samuel W Hasinoff, Paul Green, Frédo Durand, and William T Freeman, "4d frequency analysis of computational cameras for depth of field extension," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3, pp. 97, 2009.

[87] Li-Wei Kang and Chun-Shien Lu, "Distributed compressive video sensing," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 1169–1172.

[88] Aswin C Sankaranarayanan, Pavan K Turaga, Richard G Baraniuk, and Rama Chellappa, "Compressive acquisition of dynamic scenes," in *Computer Vision–ECCV 2010*. 2010, pp. 129–142, Springer.

[89] Richard G Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[90] Mahdad Hosseini Kamal, Mohammad Golbabaee, and Pierre Vandergheynst, "Light field compressive sensing in camera arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 5413–5416.

[91] Kazuki Ohashi, Keita Takahashi, and Toshiaki Fujii, "Joint estimation of high resolution images and depth maps from light field cameras," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014.

[92] Christoph Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 93–104.

[93] Amit Ashok and Mark A Neifeld, "Compressive light field imaging," in *Proc. SPIE*. International Society for Optics and Photonics, 2010, vol. 7690.

[94] S Derin Babacan, Reto Ansorge, Martin Luessi, Pablo Ruiz Matarán, Rafael Molina, and Aggelos K Katsaggelos, "Compressive light field sensing," *Image Processing, IEEE Transactions on*, vol. 21, no. 12, pp. 4746–4757, 2012.

[95] Zhimin Xu and Edmund Y Lam, "A high-resolution lightfield camera with dual-mask design," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2012.

[96] Kshitij Marwah, Gordon Wetzstein, Yosuke Bando, and Ramesh Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 46, 2013.

[97] Pierre Vandergheynst and Pascal Frossard, "Efficient image representation by anisotropic refinement in matching pursuit," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*. IEEE, 2001, vol. 3, pp. 1757–1760.

[98] Kjersti Engan, Sven Ole Aase, and J Hakon Husoy, "Method of optimal directions for frame design," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*. IEEE, 1999, vol. 5, pp. 2443–2446.

[99] Michal Aharon, Michael Elad, and Alfred Bruckstein, "k-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.

[100] Ivana Tosic and Pascal Frossard, "Dictionary learning," *Signal Processing Magazine, IEEE*, vol. 28, no. 2, pp. 27–38, 2011.

[101] Qiang Yao, Keita Takahashi, and Toshiaki Fujii, "Compressed sensing of ray space for free viewpoint image (fvi) generation," *ITE Transactions on Media Technology and Applications*, vol. 2, no. 1, pp. 23–32, 2014.

[102] Salil Tambe, Ashok Veeraraghavan, and Amit Agrawal, "Towards motion aware light field video for dynamic scenes," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1009–1016.

[103] Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 287–294.

[104] Qiang Yao, Keita Takahashi, and Toshiaki Fujii, "Overcomplete compressed sensing of ray space for generating free viewpoint images," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*. IEEE, 2013, pp. 1–6.

[105] Scott Shaobing Chen, David L Donoho, and Michael A Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[106] Emmanuel J Candès, Michael B Wakin, and Stephen P Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5–6, pp. 877–905, 2008.

[107] Dornoosh Zonoobi and Ashraf A Kassim, "On the reconstruction of sequences of sparse signals–the weighted-cs," *Journal of Visual Communication and Image Representation*, vol. 24, no. 2, pp. 196–202, 2013.

[108] "http://lightfield.stanford.edu/lfs.html/," .

Acronyms

BP	Basis Pursuit
CP	Computational Photography
CPU	Central Processing Unit
CS	Compressed Sensing
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DHT	Discrete Hadamard Transform
EPI	Epipolar Plane Image
FVI	Free Viewpoint Image
GMM	Gaussian Mixture Model
GPU	Graphic Processing Unit
GPGPU	General Purpose Graphic Processing Unit
IRLS	Iterative Reweighted Least Squares
MOD	Method of Optimal Direction
MP	Matching Pursuits
OMP	Orthogonal Matching Pursuit
PCB	Printed Circuit Board
PSNR	Peak Signal noise Ratio
RIP	Restricted Isometry Property
RMSE	Root Mean Square Error
SNR	Signal to Noise Ratio

Publications

Journal

[1] Qiang Yao, Keita Takahashi, Toshiaki Fujii, "Reconstruction of Compressively Sampled Ray Space by Statistically Weighted Model", *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences*, Vol.E97-A,No.10, (2014.10)

[2] Qiang Yao, Keita Takahashi, Toshiaki Fujii, "Compressed Sensing of Ray Space for Free Viewpoint Image (FVI) Generation", *ITE Transactions on Media Technology and Applications*, Vol. 2 (2014), No. 1, pp. 23-32, (2014.01)

International conference

[1] Qiang Yao, Keita Takahashi, Toshiaki Fujii, "Reconstruction of Compressively Sampled Ray Space by Using DCT Basis and Statistically-Weighted L_1 Norm Optimization", *IS& T/SPIE Electronic Imaging 9020, Computational Imaging XII* , (2014.02)

[2] Qiang Yao, Keita Takahashi, Toshiaki Fujii, "Overcomplete Compressed Sensing of Ray Space for Generating Free Viewpoint Images", *Proceedings of APSIPA ASC 2013*, (2013.10)

[3] Qiang Yao, Toshiaki Fujii, "Highly Efficient Representation of Ray Space for FTV System", *The 5th International Conference of 3D Systems and Applications (3DSA)*, (2013.06)

[4] Qiang Yao, Toshiaki Fujii, "Sparsity Analysis of EPI by Dictionary Learning Method", *2013 International Workshop on Advanced Image Technology (IWAIT2013)*, pp. 446-450, (2013.01)

Domestic conference

[1] Qiang Yao, Keita Takahashi, Toshiaki Fujii, "Three Dimensional Compressed Sensing of Ray Space", *Proceeding of the Tokai-Section Joint Conferences of Electrical and Related Engineers 2013*, (2013.09)

[2] Qiang Yao, Toshiaki Fujii, "Sparse Acquisition of Ray Space by Sensing EPI", *Proceeding of 3D Conf 2013*, (2013.07)

[3] Qiang Yao, Toshiaki Fujii, "An Incoherent Dictionary Learning Model for Sparse Representation of EPI", *Proceeding of the 2013 IEICE Annual Conference*, (2013.03)

[4] Qiang Yao, Toshiaki Fujii, "The Decomposition of Epipolar Plane Image by Using Dictionary Learning Method", *Proceeding of the 17th Image Media Processing Symposium (IMPS)*, (2012.10)

[5] Qiang Yao, Toshiaki Fujii, "Sparse Decomposition of EPI by Using Greedy Pursuit Algorithm", *Proceeding of the 12th Forum on Information Technology*, (2012.09)

[6] Qiang Yao, Toshiaki Fujii, "Sparse Representation of Epipolar Plane Image by Overcomplete Dictionary", *Proceeding of the Tokai-Section Joint Conferences of Electrical and Related Engineers 2012*, (2012.09)

[7] Qiang Yao, Toshiaki Fujii, "Sparse Decomposition of EPI by Greedy Pursuit Method", *The 1st WAKATE Seminar of Tokai-Section of Institutes of Electrical and Related Engineers 2012*, (2012.09)

[8] Qiang Yao, Mehrdad Panahpour Tehrani, Toshiaki Fujii, and Masayuki Tanimoto, "Overview of Multi-view Video plus Depth (MVD) Coding", *The 2nd WAKATE Seminar of Tokai-Section of Institutes of Electrical and Related Engineers 2011*, (2012.03)