Fourier Tag: A Smoothly Degradable Fiducial Marker System with Configurable Payload Capacity

Anqi Xu
School of Computer Science
McGill University
Montréal, Québec, Canada
anqixu@cim.mcgill.ca

Gregory Dudek
School of Computer Science
McGill University
Montréal, Québec, Canada
dudek@cim.mcgill.ca

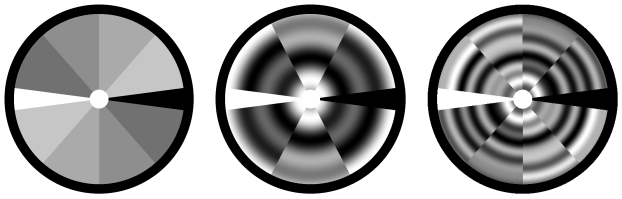
***Abstract*—We describe the design and implementation of a fiducial marker system that encodes data in the frequency spectrum of a synthetic image. This distinctive approach to marker synthesis and data encoding allows for partial data extraction in adverse imaging conditions, and can significantly extend the detection range through graceful data degradation. Additional digital encoding and image construction techniques are used to increase the payload capacity, and also to store 3-D pose information in each fiducial marker. This fiducial marker scheme can be configured to match the needs of the target application. We present several experiments investigating the practical range of various parameters as well as the performance of a specific instance of the system.**

***Keywords*-fiducial marker; graceful degradation; Fourier transform; Phase-Shift Keying.**

I. Introduction

This paper presents a new form of fiducial marker for use in vision, robotics and automation applications. Fiducial markers are artificial image-domain objects that are made to be detectable, and which usually have an information payload embedded in their graphical layout. The most familiar instance of such markers (and the least appropriate for robotics) is the room number located above office doors. The barcodes placed on commercial products also fall into this category. The "fiducial" component of such systems refers to the fact that they can also be used to infer geometric layout (i.e. the pose of the markers), although this is not used in some applications such as product labeling. The bullseye that makes up a traditional dart board is a classic fiducial symbol that emphasizes the importance of the geometric positioning application of such targets.

The construction of fiducial marker technologies reflect a tradeoff between several design factors, including most importantly the accuracy of geometric inference, payload capacity, and robustness to varying imaging conditions. There are two main classes of fiducial systems, consisting of barcode-like markers that encode relatively large amounts of payload data (e.g. $> 10^{10}$ unique IDs), and visually distinguishing symbols that have comparably smaller payload capacities (e.g. on the order of 10^3 unique IDs) but that can also provide accurate positioning information. In general, the former class of payload-based fiducials are typically used for information storage purposes, and require each marker to be captured individually and at close proximity to the camera or scanning device. On the other hand, the latter type of symbols are commonly employed as visual landmarks whose ID, location and possibly pose can be extracted reliably even from images containing multiple markers.



(a) Layout (b) ID=6 (6 bits) (c) ID=625518 (36 bits)

Figure 1. Fourier Tag is a family of planar fiducial markers that encodes binary payload data into a grayscale, radially symmetric structure (a). (b) shows a tag with 3 symmetric arc sector pairs, each containing 2 bits of payload and 1 bit of checksum, whereas (c) depicts a tag with 4 data sections, each containing 9 bits of payload and 3 bits of checksum.

An intrinsic challenge with fiducial mechanisms is that as the distance between the imaging device and the marker increases, or as viewing conditions degrade, the ability to extract the *payload* (i.e. digital content stored in the marker) degrades. For most existing systems this degradation occurs as a step function going from full payload to nil, even when error control codes are used to extend the detection range. The sudden change from full to no data extraction is known as abrupt degradation, and is undesirable in virtually all applications. While it may seem inevitable, this is rarely the case with human perception and classification: as a person moves away from us, there is a point where we may no longer recognize their face yet we can still determine their identity, and at much greater distances, we can still perceive that they are human.

This paper presents a fiducial marker scheme that is designed specifically to address the abrupt data degradation issue. This is achieved by encoding data in the frequency domain to take advantage of the classical phenomenon whereby content in lower frequencies are more robust to many sources of image degradation than higher frequencies [1].

978-0-7695-4362-8/11 $26.00 © 2011 IEEE
DOI 10.1109/CRV.2011.13

The fiducial symbols we have developed, referred to as "Fourier Tags" and shown in Fig. 1, allow part of their payload to be extracted even when the marker size is small or when the image is blurry. This graceful degradation effect is achieved by storing data as both low-frequency and high-frequency sinusoidal patterns. As imaging conditions worsen, the content stored in lower frequencies will remain detectable even after the data in higher frequencies has been destroyed by noise.

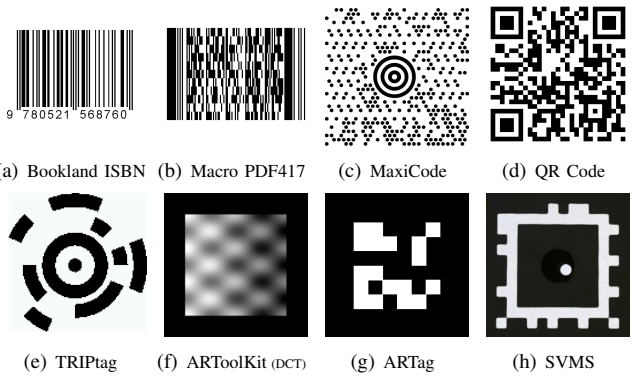
Developers can take advantage of this unique property by constructing the payload in such a way that knowing part of the data can still be meaningful. For example, consider a Fourier Tag that embeds both specific radioactive properties of a hazardous canister as well as a general radioactive warning. Assuming that the general warning component can be detected from afar, then pedestrians will be able to extract this useful information without having to be in physical proximity to the marker. On the other hand, properly suited professionals can take a close-up image of the tag to reveal technical data, which can help them decide on how to properly handle the radioactive contents.

Fourier Tag's encoding process contains several adjustable parameters that determine the visual shape of the marker as well as the size of the payload. By tuning these parameters based on the needs and conditions for a particular application, we can create an instance of the Fourier Tag scheme which reduces unused payload size to increase the robustness of the image detection process and the encoding strength. Since Fourier Tags can also be used to extract pose information, this configurable family of graceful degradable fiducial markers exhibits benefits from both payload-based and landmark-based fiducial systems.

In this paper, we extend the prior work on the Fourier Tag concept [2] by presenting a practical and usable realization, a different and significantly improved data encoding scheme with potentially more than two orders of magnitude increase in payload capacities, and a concrete implementation for detecting Fourier Tags in images. We also we describe an evaluation methodology for tuning various system parameters to adapt the Fourier Tag scheme for a specific deployment. This is illustrated with a number of empirical studies for determining practical limits of various system parameters. We further present a cursory performance assessment of a specific instance of this technology.

II. Related Work

Among existing 2-D fiducial marker systems, payload-based symbols have been refined extensively and are widely used in commercial applications, whereas landmark-based tags are largely confined to research and exotic applications. The first row in Fig. 2 illustrates several payload-based fiducial systems that are well known to the mass public. These include the Universal Product Code (UPC) family of barcodes (Figs. 2(a) and 2(b)), used to identify merchandise;



(a) Bookland ISBN (b) Macro PDF417 (c) MaxiCode (d) QR Code

(e) TRIPtag (f) ARToolKit (DCT) (g) ARTag (h) SVMS

Figure 2. Examples of various fiducial marker systems.

the Maxicode (Fig. 2(c)), employed by the United Parcel Service to track packages; and the QR code (Fig. 2(d)), which can embed URLs and other textual content.

The second row in Fig. 2 shows several landmark-based fiducial marker systems sharing similarities to the Fourier Tag scheme. The TRIP (Target Recognition using Image Processing) infrastructure [3] comprises of a circular fiducial system called the TRIPtag (Fig. 2(e)). These bi-tonal markers are detected in images by scanning for their bullseye structure comprising of a solid ring and a central dot. Two concentric outer rings are individually shaded in each of its 16 arc sectors. A ternary system is used to encode $3^9 = 19683$ unique IDs along with a parity checksum. In addition, the physical radius of each generated tag is also encoded in its structure, and can be used along with its circular shape to compute the marker's full 3-D pose in the camera's frame. The downside however is that these tags cannot be rescaled without modifying its contents. TRIPtags also cannot handle occlusion, since the stored checksum can only detect errors but not correct them.

ARToolKit [4] and ARTag [5] are two similar types of square-shaped fiducial marker systems, both conceived originally for Augmented Reality applications (hence the *AR* prefix). These fiducials are identified by their square contour, which is used to compute the pose of the marker in the camera's frame. ARToolKit markers are constructed using arbitrary monochrome or grayscale patterns (Fig. 2(f)), whereas ARTag symbols are structured as a lattice of bi-tonal squares (Fig. 2(g)). During detection, both systems sample the square visual payload into a description matrix, and whereas ARToolKit matches the 16×16 sampled matrix to a variable sized pattern database using numerical correlation, ARTag employs digital error correction techniques to reliably extract 10 bits of payload data from the sampled 6×6 intensity matrix. Although in principle ARToolKit can recognize more unique symbols than ARTag, this advantage can be hard to exploit in practice without risking inter-symbol confusion. In contrast, the 2002 unique markers recognizable by the ARTag system were selected to specifically minimize these error rates. The DCT variant of

ARToolKit [6] is also designed to help reduce ARToolKit's inter-marker confusion rates, by encoding data as distinct amplitudes in the frequency spectrum using On-Off Keying (OOK) modulation [7]. Although the resulting patterns are maximally distinct from each other, this ARToolKit variant is limited to 256 unique patterns.

The Space Vision Marker System (SVMS) [8] consists of bi-tonal fiducials with a solid white square contour surrounded by 44 black or white cells encoding 15 bits of data (Fig. 2(h)). Similar to ARToolKit and ARTag systems, the location and pose of SVMS markers are detected from their square contours. At the center of each SVMS marker lies a protruded black cylinder with a concentric white dot. When viewing this 3-D structure from an angle, the white lid's off-centered position can be used to refine the pose of the target symbol. SVMS boasts both a relatively large payload capacity as well as highly accurate pose estimates, although the manufacturing of the 3-D structure may decrease its appeal in applications requiring cheap and fast deployment.

III. Methodology

The layout of the latest incarnation of the Fourier Tag fiducial technology is illustrated in Fig. 1(a); each tag is characterized by its concentric black outer ring and white inner circle. The area between these two elements exhibits visual regularities to facilitate detection, and is partitioned into an even number of arc sectors. Two special *alignment sectors* with fixed arc widths, colored in solid black and white respectively, are used to indicate the orientation for each fiducial. The payload data is encoded into S signals, where each signal is constructed from a sum of sinusoids. These signals are plotted as grayscale patterns within S symmetric pairs of data sectors along the radius of the entire circular structure. This design ensures that most of the Fourier Tag structure is radially symmetric, with the exception of its two alignment sectors.

We describe our formulation of the Fourier Tag scheme as three algorithmic components: the encoding of information into visual-domain symbols, the detection of markers within an acquired image, and the decoding and validation of the payload content.

A. Fourier Tag Synthesis: Partitions, Sections and Capacity

Each Fourier Tag stores a digital payload that can be interpreted as the tag's numerical identifier. This payload can be partitioned arbitrarily into different components, and the synthesis process can be configured correspondingly to be able to individually validate the integrity of each partition. Because these payload components are embedded in the frequency domain within distinct frequency ranges, when a marker is observed in an image, the system will be able to return payload partitions stored in lower frequencies even when image-domain noise (e.g. blur, positional quantization, pixel intensity noise) has destroyed high frequency content.

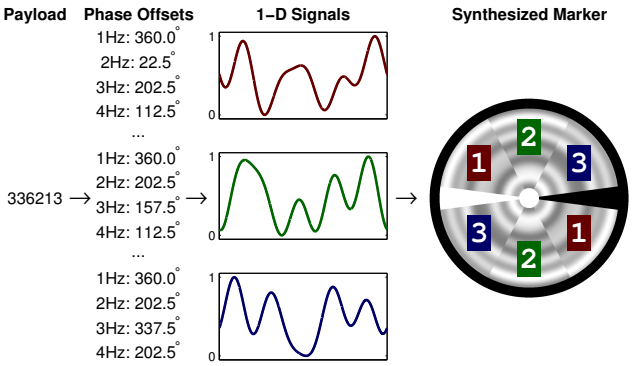


Figure 3. Illustration of the Fourier Tag synthesis process: a numerical payload is separated into $S = 3$ data sections; each section is encoded into a set of phase offsets at various frequencies of a 1-D signal; these signals are plotted as grayscale patterns into S symmetric pairs of arc sectors.

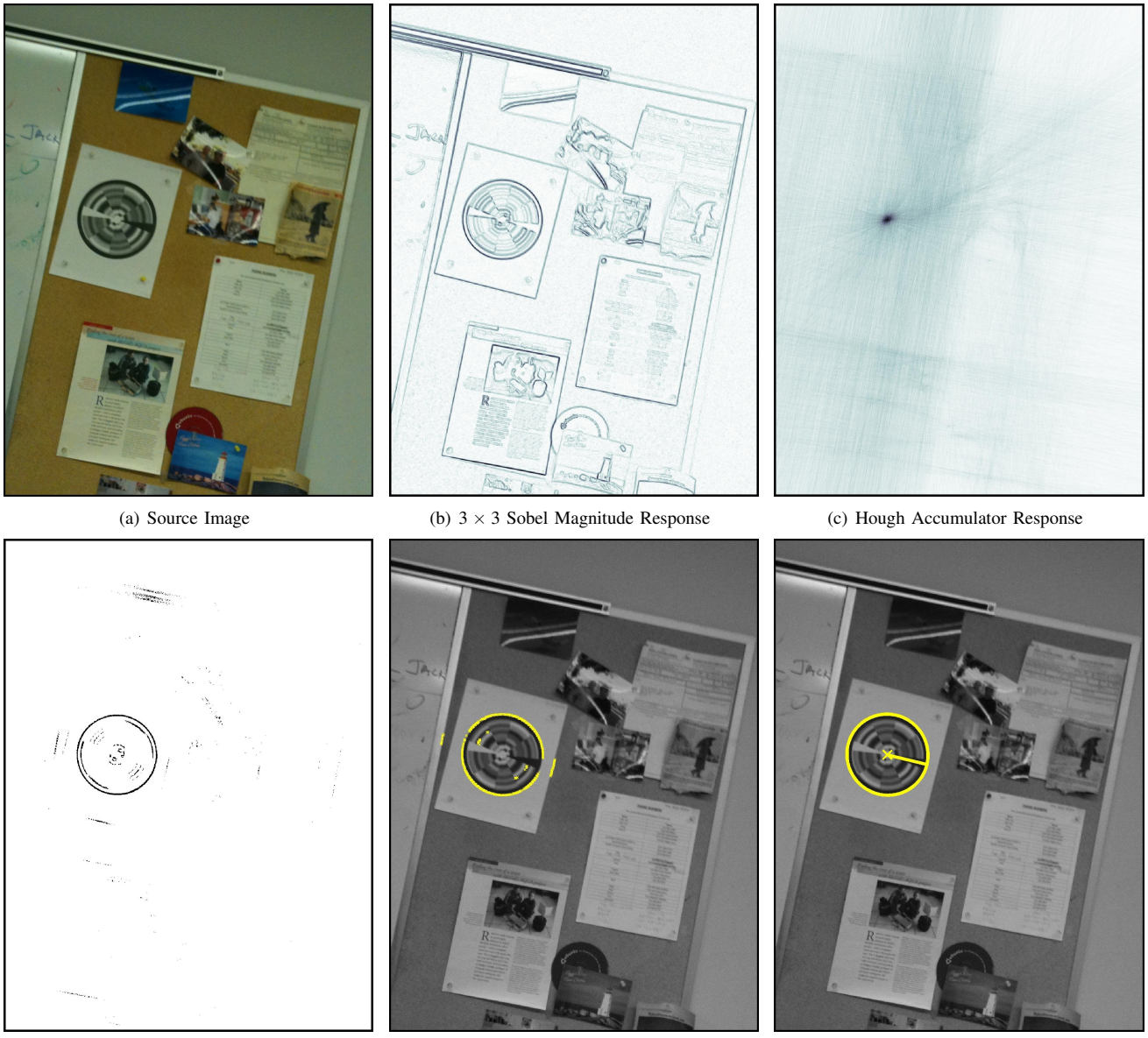
The entire payload is broken down into S *sections*, by uniformly distributing content from each partition into every section. These sections correspond to the S pairs of arc sectors in the Fourier Tag layout in Fig. 1(a), where every two radially opposing arc sectors both encode the same data. Each payload section is remapped into an "augmented representation" to enhance robustness to image-domain noise and to be able to detect any resulting data corruption. The specifics of this encoding function is described later.

The data in each encoded payload section is then stored in the frequency domain of a 1-D intensity pattern using Phase Shift Keying (PSK) modulation [7], which is also used in communication technologies such as WiFi and Bluetooth. PSK works by breaking up the data into many small portions and storing each portion as the phase offset of a unique sinusoidal component (i.e. frequency). Specifically, if each data portion can take on one of M values, then the data to be stored can be expressed in a base-M representation m_i, for $i = 1, 2, ..., N$. Assuming that M is a power of 2 ($M = 2^B$), then each digit m_i stores B bits of information. Next, each digit is scaled into an angular value and is used to shift the starting phase of a selected sine wave at frequency $f_i = i + f_0$, where f_0 is a fixed offset. All the sine waves are then superimposed together, and the final result is obtained by sampling one period of this continuous signal into a T-length vector:

$$psk[t] = \sum_{i=1}^{N} sin\left(2\pi f_i \cdot \frac{t}{T} + \frac{2\pi}{M} m_i\right) \qquad (1)$$

where $t = 1, 2, ..., T$.

Assuming that the phase modulation factor is a power of 2 ($M = 2^B$), the total binary payload capacity for a particular configuration of the Fourier Tag scheme can be computed as $C = B$ (bits per phase channel) $\times$ N (phase channels per data section) $\times$ S (data sections). We can thus adjust B, N and S to determine an appropriate Fourier Tag configuration for the specific payload needs of a target application.



(a) Source Image (b) 3×3 Sobel Magnitude Response (c) Hough Accumulator Response

(d) Potential Marker Edgels (e) Potential Marker Border Pixels (f) Ellipse Fit and Orientation Extraction

Figure 4. Illustration of the steps in the detection process for finding Fourier Tag structures in a given image.

Naturally, signal vectors extracted from Fourier Tags observed in images will contain various types of noise. As result, the recovered phase angles will generally be shifted by a certain amount from its original values. Because PSK encodes consecutive values in base-M into phase offsets that differ by $\frac{2\pi}{M}$, the recovered phase can be re-mapped into the correct base-M digit as long as the magnitude of the error remains below $\frac{1}{2}\frac{2\pi}{M}$. If the error in the recovered phase is larger than this threshold, then it will be mapped onto an incorrect base-M value. To be able to detect these data errors, during synthesis the system converts each data portion m_i to a B-bit gray code, and then appends extra data portions containing error checksums.

The size and format of the error checksum should be configured to match the expected level of image-domain noise for the target application environment. For example, a 1-bit parity checksum can identify a single error caused by a phase offset difference of up to $\pm\frac{2\pi}{M}$. A checksum must be generated for each payload partition, to be able to validate their data consistency independently.

In addition to storing data in its structure, each Fourier Tag also contains a signature pattern, which minimizes the likelihood of falsely detecting these patterns in natural scenes. This is accomplished by injecting fixed amplitude and phase values in the frequency domain below the data frequency range f_0. In the majority of applications it is

sufficient to use the simplest signature configuration, which is to set $f_0 = 2\ Hz$ and append a $1\ Hz$ sine wave with zero phase to each signal, i.e. add $sin\left(2\pi\frac{t}{T}\right)$ to the right hand side of Equation 1.

The 1-D signal for each of the S arc sections is rescaled as grayscale intensity values and plotted radially outwards in its respective pair of arc sectors, as depicted by Fig. 3. Sub-pixel stratified sampling [9] is used to minimize the appearance of visual aliasing artifacts.

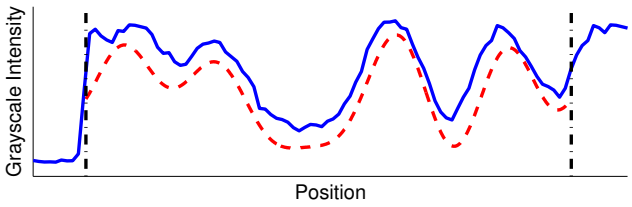
B. Fourier Tag Detection

An automated detection process identifies the Fourier Tag structure in a given image by looking for its characteristic elliptical and radially symmetric shape, as illustrated by Fig. 4. This process involves carrying out edge detection, estimating the tag center using both a voting scheme and a symmetry operator, and determining the marker's contours first as pixel positions and subsequently as an elliptical fit. This detection approach can find a Fourier Tag marker even if it is partially slanted away from the camera or is partially occluded. Our description here considers only the detection of a single Fourier Tag per image, and the case of multiple tags per image is a natural extension.

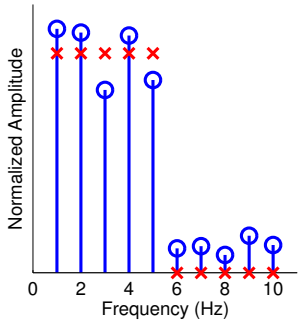
First, an edge detection algorithm is used to generate gradient magnitude and direction maps of the given grayscale source image. Since the result does not need to be extremely accurate at this stage, the Sobel operator is chosen for its simplicity and speed. Next, a technique inspired by the circular Hough transform [10] is used to detect the center of the marker. For each pixel in the image, a line is traced along the direction of the estimated intensity gradient. Each line is weighted by the gradient magnitude of the source pixel and is collected into a so-called "Hough accumulator map", as illustrated in Fig. 4(c). An estimate of the fiducial's center is computed by thresholding the accumulator response and obtaining the centroid position of the strongest cluster. This estimate is subsequently refined by scanning within a local window of the grayscale source image for the location with the strongest amount of radial symmetry.

The system then determines the border of the Fourier Tag structure, by first applying an adaptive threshold to the gradient magnitude response. The resulting map is filtered by selecting edge elements whose gradient directions are approximately aligned with the estimated center point (Fig. 4(d)). Border positions are obtained by scanning radially outwards from the center point at various angles and searching for entries belonging to a large connected group. The analytical form of the elliptical marker structure is computed using the RANSAC algorithm [11] applied to a least mean square ellipse fitting procedure [12]. The center of the ellipse is in general different from the center of the planar marker as a result of foreshortening.

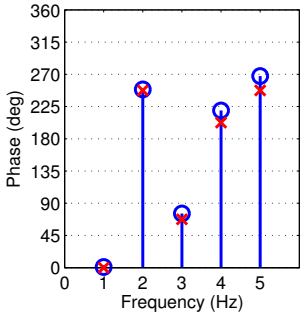
To obtain pose information of the detected marker, the system extracts multiple *rays* by sampling the grayscale source image starting from the marker's center point and ending at its elliptical boundary, at uniformly spaced angles. Since the alignment sectors of any Fourier Tag always contains a single color, intensity vectors lying within these sectors will have smaller variance compared to rays on top of data sectors. The planar orientation of the marker is therefore determined by convolving the variance in grayscale intensity for each extracted ray with a binary mask highlighting the angular positions of the alignment sectors, and finding the position with the minimum variance in the convolved response. Once the in-plane orientation is found, 3-D pose of the marker can be deduced from the geometry of the ellipse, using the approach outlined in [3]. In particular, the positional difference between the ellipse center and the marker's symmetric center point is used to resolve a projection-related ambiguity. Unless the physical radius of the marker is known beforehand or is stored as part of the payload, the distance between the camera and the marker will be off by an unknown scalar factor.



(a) Extracted Ray (solid line) / Superimposed Original Signal (dashed line)



(b) Amplitude Spectrum



(c) Phase Spectrum

Figure 5. In (a), a ray extracted from an observed Fourier Tag (solid line) is trimmed from both extremities to remove marker boundary content (at dash-dotted lines). The trimmed ray is generally not identical to the original signal (dashed line). The amplitude spectrum (b) and phase spectrum (c) responses (circles) also differ from their original values (cross), although the binary payload will be unaffected as long as the observed phase angles lie within the same grid sector as their original values.

The modulated signal for each of the S data sections is recovered by retrieving rays whose angular offsets from the marker's orientation fall within the respective arc sector. To remove the border values corresponding to Fourier Tag's concentric white dot and black ring respectively, each signal is trimmed from both extremities until their pixel intensities converge. An example of the original ray and the indexes generated the trimming algorithm are shown in Fig. 5(a).

C. Fourier Tag Decoder

All the extracted and trimmed rays are individually processed through the Fast Fourier Transform (FFT) to reveal their frequency-domain contents. Although individual intensity values of the observed Fourier Tag can be perturbed by noise from the marker synthesis, printing, imaging and detection processes, a valid ray will retain the general shape of its original signal, as shown in Fig. 5(a). Similarly, both the amplitude spectrum and phase spectrum responses will differ numerically from the original encoding, although the effects of these errors are mitigated by the encoding scheme.

The presence of the Fourier Tag's signature is verified by comparing the relative amplitudes and absolute phase values at the injected frequencies. For example, to identify the signature recommended in Sec. III-A (i.e. $phase_{1Hz} = 0^{\circ}$), one should look for the presence of a strong amplitude and a near-zero phase at 1 Hz frequency. If either condition is not satisfied, then the signal is discarded.

Once the signature of the Fourier Tag has been validated, the phase spectrum entries are re-scaled back into base-M (or B-bit) strings of digits. After separating the checksum portions, gray coding is used to decode the rest of the data. Each payload partition is validated against its respective checksum to ensure that the recovered data is consistent. For safety, all payload portions stored in higher frequencies than an invalidated portion are discarded due to the low-pass nature of various types of image-domain noise.

Because each of the S data sections are stored in two radially opposing arc sectors within the Fourier Tag layout, the encoded information can still be retrieved even if one of the two sectors is occluded in the image. In general, multiple rays are extracted from each data section, and after applying the decoding process described above to each ray, a consensus scheme is applied to all copies of each payload portion that passed their checksum. This precautionary step is used to detect errors missed by the checksum.

D. Sources of Error

The amount of payload data successfully extractable from an observed Fourier Tag marker can be affected by many sources of error. Although some factors, such as the size and slant of the marker in the image, are inherent to the setup and are not always changeable, this section discusses general strategies for ensuring maximal data integrity within the ray vectors extracted from an image.

Whether the tag is printed or displayed digitally, care should be taken to maximize both the spatial resolution and grayscale resolution of the result. This will respectively minimize the presence of aliasing and quantization noise.

The imaging device should be configured to maximize the contrast sensitivity of the captured fiducial marker. This minimizes the quantization on the extracted pixel grayscale intensity values. Ambient lighting does not affect frequency content beyond a DC offset, although an under- or over-exposed image can cause pixel saturation, which will mostly damage high-frequency content. An out of focus marker will also exhibit more error in higher frequency contents.

Inaccuracies in both locating the marker center and the elliptical boundary will negatively impact the positional accuracy of extracted rays. If both the starting and ending positions of a ray are shifted by a same amount along the radius of the marker, this will add an offset to each response in the phase spectrum proportional to its carrier frequency. On the other hand, if the system does not extract the entire length of a stored signal, the frequency axis for both the amplitude and phase spectra will be scaled by a non-integer amount. This so-called "frequency leakage" effect causes a blurring-like effect of the amplitude and phase values into adjacent frequencies. Any other inaccuracies in the ray extraction process will affect the frequency spectrum in complex and often unpredictable ways.

IV. Experimental Evaluation

In this section we present several experiments conducted in real environments to estimate key parameters of our Fourier Tag scheme. We also describe how the system can be tuned for specific applications by presenting an illustrative sample configuration. In general, the performance of a Fourier Tag configuration depends on the viewing distance and angle, camera parameters, printer quality, and other variables. We believe that our sample instance is indicative of the system performance of this general approach, and can be varied easily for a different set of conditions.

As explained in Sec. III-A, the total payload capacity C of a Fourier Tag configuration depends on the number of data sections S, the number of phase channels (i.e. carrier frequencies) needed per data section, N, and the number of bits of data encoded into each phase channel B. There are a number of secondary implementation considerations as well, such as the payload partitioning and error checksum selection, although they are outside the scope of what we can describe in this paper.

The two experiments presented below determine practical upper bounds on the number of phase channels allowed, N, and on the number of bits of data stored into each phase channel, B (assuming that $M = 2^B$). In these experiments, rather than using PSK the signal is artificially generated by injecting random phase offsets into every possible carrier frequency, from 1 Hz up to a maximum value f_{max}. Each phase offset is assumed to represent a base-M digit. The artificial payload of the generated signal is assumed to be separated into 2 partitions, corresponding to a *primary payload component* stored at low frequencies and a *secondary payload component* stored at high frequencies.

The synthesized signals are embedded into the Fourier Tag structure, which is then displayed on a LCD monitor at a physical tag radius of $10\,cm$. This introduces less signal distortion compared to a printing process, as in practice additional work will be required to configure and calibrate

Table I
PAYLOAD CAPACITY ANALYSIS FOR PSK SIGNALS GENERATED USING VARIOUS FREQUENCY BANDWIDTH VALUES f_{max}

$\mathbf{f_{max}}$	Mean μ	Std. Dev. σ	3σ Margin	$\mathbf{M_{max}}$	Capacity
$5Hz$	5.8274°	3.9904°	17.7986°	10	16 bits
$10Hz$	7.2564°	5.4719°	23.6721°	7	28 bits
$15Hz$	13.6803°	9.9997°	43.6794°	4	30 bits
$20Hz$	18.3552°	12.8602°	56.9358°	3	31 bits

the printer to produce results with optimal contrast and preserved linearity. After taking images of markers at various distances, the center point and boundary locations for each marker are manually entered to remove error introduced by the automated marker detection process.

The experiments were conducted using a 2 megapixel 640×480 USB "webcam", given its popularity in mass market electronics such as laptops and cellphones. To represent conditions of a typical indoor usage scenario, the objective in both experiments is to recover the encoded data from both payload components at a distance of $0.5m$ (i.e. 150 pixel marker radius) and from at least the primary component at a distance of $3m$ (i.e. 25 pixel marker radius). Bounds on N and B are chosen to meet our performance objective, and the second experiment also determines optimal frequency ranges for storing both payload components.

A. Frequency Bandwidth Experiment

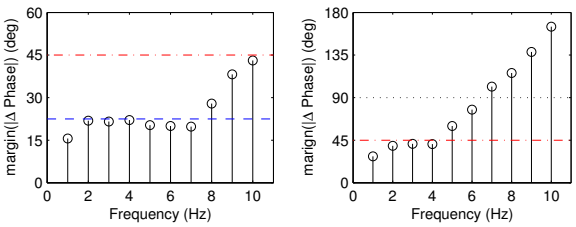
The frequency bandwidth f_{max} of a signal is the highest frequency in which data is stored. We assume that data is injected into every available integer frequency to maximize payload capacity. Because all the carrier sine waves are superimposed (see Equation 1) and quantized into a grayscale range, the number of sinusoids will be proportional to an increased need in resolution. Thus the frequency bandwidth and subsequently the number of phase channels available (N) are ultimately limited by the image's grayscale resolution.

We synthesized multiple signals for four different bandwidth settings. The resulting fiducial markers are displayed on a LCD screen at a physical radius of 10 cm, whereas a camera directly facing the monitor is placed at 0.5 m distance. Images are captured in a well-lit indoor office environment. The extracted phase values are compared to their original angles to measure the magnitude of the noise at all frequencies in the phase spectrum. We define the tolerance margin as an expected upper bound on the noise, i.e. $margin = \mu + 3\sigma$. The binary payload capacity per signal can be estimated as:

$$capacity = \lfloor f_{max} \cdot \log_2(M_{max}) \rfloor \quad (2)$$

where $M_{max} = \lfloor \frac{1}{2} \cdot \frac{2\pi}{margin} \rfloor$ represents the maximum base of the data stored in each phase channel.

As shown in Table I, the payload capacity nearly doubles when f_{max} changes from 5 Hz to 10 Hz, even though a smaller modulation basis M_{max} is required for $f_{max} = 10$ Hz to accommodate the increased error in the high-frequency phase channels. The additional increase in payload capacity



(a) $0.5m$ distance from marker (b) $3m$ distance from marker

Figure 6. Expected 3σ upper margins of error in the phase spectrum with $F_{max} = 10$ Hz for 10 cm fiducial markers observed from various distances.

is less significant for larger values of f_{max}. Given these observations, $f_{max} = 10$ Hz appears to strike an acceptable compromise between the amount of data per phase channel and the maximum expected error level in the phase spectrum.

The capacity analysis above assumes that phase channels are only used to store payload data. In practice, because the Fourier Tag encoding scheme also stores error checksums and signature content in these phase channels, the number of phase channels available for the payload of an actual Fourier Tag marker will be bounded by $N \leq f_{max} - f_0$.

B. Payload Partitioning Experiment

In this experiment our objective is to determine a reasonable partitioning of the phase spectrum for storing the primary and secondary payload partitions, as well as the storage capacity for each phase channel, B. Assuming a fixed frequency bandwidth of $f_{max} = 10$ Hz, we used the same setup as in the previous experiment, and collected images of each displayed Fourier Tag marker taken from distances of $0.5m$ and $3m$. After computing the errors in the phase spectrum, the results are aggregated into 3σ tolerance margins for the error magnitude in each frequency.

Fig. 6 illustrates that high-frequency content contains more noise than low-frequency content as result of various image-domain noise. The linear shape of the error response at 3 m is also partially caused by positional inaccuracy of each sampled 25 pixel ray. The dashed and dash-dotted lines in Fig. 6(a) represent thresholds for the storage capacity at $B = 3$ bits and $B = 2$ bits, respectively, and similarly the dash-dotted and dotted lines in Fig. 6(b) represent thresholds for data capacities of $B = 2$ bits and $B = 1$ bit. The error levels in all 10 phase channels fall below the $B = 2$ threshold when the camera is placed at $0.5m$ away from the fiducial marker. On the other hand, the error in the $f = 5 - 10$ Hz range grows beyond this threshold as the camera is moved away from the marker to a distance of $3m$. In fact, the error at frequencies higher or equal to $f = 6$ Hz can no longer reliably encode any information at all at $3m$.

Given these observations, a reasonable allocation strategy is to place the primary payload component in the $f = 1$ to 4 Hz range and store the secondary payload component in the $f = 5$ to 10 Hz range. Under this partitioning, we can reliably encode $B = 2$ bits of information in all of the 10

frequency values. Such a data allocation ensures that the content stored in the primary payload component will be unaffected by errors in the phase offsets, even when the fiducial marker is viewed from afar. The payload capacity can be further increased by using more payload partitions and frequency-specific, non-binary modulation factors (M), which we leave as an exercise for the reader.

C. System Configuration Assessment

Based on the results obtained above, we configured an instance of the Fourier Tag system to be used in indoor environments suitable for our robotics applications. We divided 36 bits of payload into 3 partitions, distributed evenly among $S = 4$ data sectors, where each phase channel contains $B = 3$ bits of data for the $N = 3$ frequencies at 3 to 5 Hz. We also stored 3 1-bit parity checksums at $f = 2$ Hz in each signal, and used our recommended signature by storing a phase value of 0° at $f = 1$ Hz. Fig. 1(c) shows a sample marker generated using this configuration, and Fig. 5 illustrates the general layout of both the amplitude and phase spectra.

Using a 5-megapixel digital camera to take images of printed Fourier Tags with a physical radius of $10\,cm$, we observed that all 3 payload partitions could be extracted reliably at distances of up to $3\,m$. As we moved further, various arc sectors began returning payloads that were inconsistent with their checksums. By manually extracting rays from images taken beyond $6\,m$ away from the marker, we remained able to extract data successfully from all 3 frequencies. We were even able to retrieve the payload stored at $f = 3$ Hz from a Fourier Tag with a 16-pixel radius. Due to the large number of variables and concerns involved, a complete study of the payload degradation as a function of imaging conditions is beyond the scope of this paper.

V. Conclusion

In this paper we have described an elaboration of the original Fourier Tag concept [2] to create practical fiducial markers whose payload capacity degrades gracefully with increasing image-domain noise. We presented a robust data encoding technique based on storing digital content into the phase spectrum of 1-D grayscale signals. Multiple signals are interleaved and stored within a circular layout, thereby increasing the total payload capacity. This general marker synthesis approach can be configured in a multitude of ways to accommodate an arbitrary payload, containing multiple partitions, each with variable size. We also described a software implementation for locating these tags in images and for extracting both the stored payload as well as pose information.

The Fourier Tag scheme can be parametrically configured to achieve a specific tradeoff point between robustness and payload capacity. We presented two experiments illustrating a general approach to performance characterization that allows the parameters of this system to be estimated. The results of these experiments were used to determine upper bounds for different parameters under a specific setting.

We are conducting ongoing studies to assess the robustness of the marker detection scheme under rotational and lighting variance, and for different display and imaging technologies. We are also hoping to conduct in the near future a performance comparison between different Fourier Tag instances and against existing landmark-based fiducial systems. Finally, we are interested in further optimizing the data encoding scheme to be able to store payload capacities competitive or beyond those of commercial fiducial systems.

References

[1] H. von Helmholtz, *Treatise On Physiological Optics*. Rochester, NY, USA: Optical Society of America (Southall JPC, ed.), 1924, vol. 2.

[2] J. Sattar, E. Bourque, P. Giguere, and G. Dudek, "Fourier tags: Smoothly degradable fiducial markers for use in human-robot interaction," in *4th Canadian Conference on Computer and Robot Vision (CRV '07)*, Montréal, Québec, Canada, May 2007, pp. 165–174.

[3] D. L. de Ipiña, P. R. S. Mendonça, and A. Hopper, "TRIP: A low-cost vision-based location system for ubiquitous computing," *Personal and Ubiquitous Computing*, vol. 6, pp. 206–219, January 2002.

[4] I. Poupyrev, H. Kato, and M. Billinghurst, *ARToolKit User Manual Version 2.33*, Human Interface Technology Lab, University of Washington, Seattle, Washington, 2000.

[5] M. Fiala, "ARTag, a fiducial marker system using digital techniques," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Washington, DC, USA, 2005, pp. 590–596.

[6] C. B. Owen, X. Fan, and P. Middlin, "What is the best fiducial?" in *Augmented Reality Toolkit, The First IEEE International Workshop*, 2002, pp. 98–105.

[7] M. Weik, *Computer Science and Communications Dictionary*. Springer, 2000, vol. 1.

[8] M. Bondy, R. Krishnasamy, D. Crymble, and P. Jasiobedzki, "Space vision marker system (SVMS)," in *AIAA SPACE 2007 Conference & Exposition*, September 2007.

[9] M. Pharr and G. Humphreys, *Physically Based Rendering, Second Edition: From Theory To Implementation*, 2nd ed. Morgan Kaufmann, July 2010.

[10] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[12] A. W. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, May 1999.