Fourier tags: Smoothly degradable fiducial markers for use in human-robot interaction

Junaed Sattar
Computer Science
McGill University
Montréal, Québec

Eric Bourque
DIRO
Université de Montréal
Montréal, Québec

Philippe Giguère
Computer Science
McGill University
Montréal, Québec

Gregory Dudek
Computer Science
McGill University
Montréal, Québec

Abstract

In this paper we introduce the Fourier tag, a synthetic fiducial marker used to visually encode information and provide controllable positioning. The Fourier tag is a synthetic target akin to a bar-code that specifies multi-bit information which can be efficiently and robustly detected in an image. Moreover, the Fourier tag has the beneficial property that the bit string it encodes has variable length as a function of the distance between the camera and the target. This follows from the fact that the effective resolution decreases as an effect of perspective. This paper introduces the Fourier tag, describes its design, and illustrates its properties experimentally.

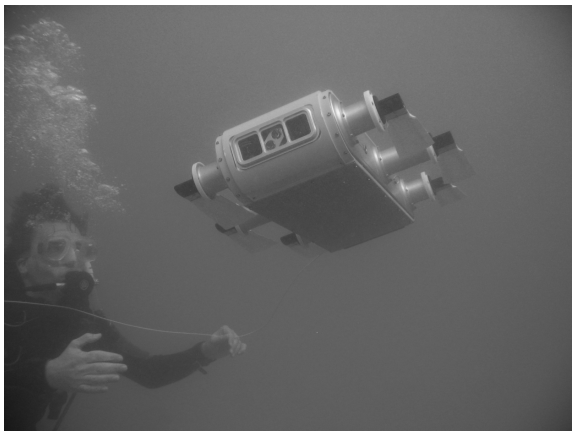


Figure 1. An underwater robot with a diver.

1 Introduction

Fiducial markers are used to provide visual cues that are easy to detect for computer vision systems. They are used to provide reliable ground truth position estimates for robotics applications, and can be used to label individual real-world objects for use in manipulation or scene augmentation. The Fourier tag gains its name by virtue of the fact that its design is based on encoding a bit pattern explicitly in the amplitude spectrum of the Fourier transform of the tag. Our work with the Fourier tag was inspired by our prior use of ARTag markers [5], an alternative set of robust fiducial markers designed for use in virtual reality applications. The ARTag markers, and all similar fiducial systems we are aware of, encode one or more bits of data (the information payload) using a geometric pattern. If the viewing conditions deteriorate (due to distance, camera noise, fog, or other factors) the pattern eventually becomes ambiguous and no further information is extractable. Note that some systems such as ARTag use error correcting codes which can tolerate partial occlusion, but eventually even such robust systems fail. In the applications we have considered, and specifically in using tags for mobile robot control, we have observed that circumstances often arise where a landmark is observable, but its pattern information is not clear enough to be used.

The Fourier tag was developed[1] to address precisely the problem of inadequate resolution, as well as simple detection and fast decoding. Specifically, as a Fourier tag is viewed with diminishing accuracy, the information it encodes degrades gracefully. When partly recognized, the high-order bits of the numerical encoding of the pattern's identity are preserved, and the low-order bits decay away successively. This is because the Fourier tag encodes bits in the amplitude spectrum in the frequency domain, with successively lower-order bits using successively higher frequencies. Since the imaging process can be approximated as a low-pass filtering transformation of the image, the low-order bits are selectively lost as the image of the tag loses resolution with distance (due to many processes that can include perspective foreshortening and atmospheric scattering).

[1]The first validation of the concept was carried out by Omar Abdul-Baki under the supervision of Dudek.

Fourth Canadian Conference on Computer and Robot Vision(CRV'07)
0-7695-2786-8/07 $20.00 © 2007 IEEE

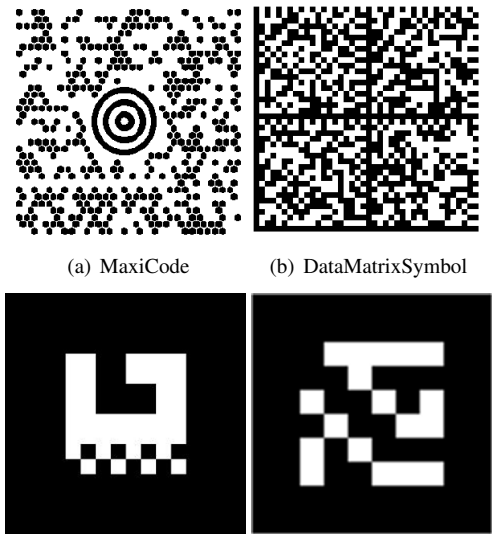
One of the problems we are currently investigating [3] is that of a diver instructing a swimming semi-autonomous robot underwater, giving it instructions so that it can carry out certain maneuvers and observations and also possibly report back to the diver as shown in Fig. 1. Using radio communications underwater is not feasible, since the energy of the radio wave is rapidly absorbed by salt water. Using a tether under water is cumbersome, and acoustic modems have limited bandwidth, consume more power, and are more susceptible to noise. Scuba divers commonly resort to hand signals to send messages to each other to coordinate tasks underwater, and in that spirit, our vehicle is equipped with an operating mechanism based on visual cues and target following.

Currently, our system uses ARTags as visual cues to communicate with the robot. With the application of Fourier tags, this communication between the divers and the robot promises to become more robust, since the distance up to which a tag can be identified is improved significantly. While the uniqueness of the tag might be a question, the fact that a visual pattern can be identified as a fiducial marker from farther out certainly promises to be an advantage for underwater visual communication.

2 Background

Systems that use fiducial markers can be decomposed into an operational chain involving three phases: 1) the design and synthesis of the fiducial itself, 2) the detection and extraction of the fiducial in the image, and 3) the decoding of the payload from the marker image. In many systems the fiducials carry little semantic information, and may not even be labelled so the third stage is missing. Indistinguishable light emitting diodes are an example of markers of this type [13].

A number of planar marker systems exist that embed information in encoded patterns, with a few designed explicitly for localization as required for applications like augmented reality (AR) [9]. Among some general purpose marker schemes, the US Postal Service uses *MaxiCode* (Fig. 2(a)) to encode shipping information in packages, while *QuickResponse* and *Data Matrix* (Fig. 2(b)) are also used for containing information needed for part labeling. These encoding methods all use error correction schemes to recover the encoded information in cases where part of the information retrieved from the symbols can be corrupted. Although useful for encoding information with more sophistication than standard commercial bar-codes, these systems are not very useful as fiducials. There are two reasons for this: first, they can be hard to detect in a large field-of-view with perspective distortion and second, they have to occupy a large portion of the image to be detected robustly. The latter requirement severely limits the range at which



(a) MaxiCode (b) DataMatrixSymbol

(c) ARToolkit (d) ARTag

Figure 2. Different existing planar markers.

these markers can be detected.

ARToolKit [7] and ARTag markers (Figs. 2(c) and 2(d) respectively) are designed to be used as fiducial markers for augmented reality applications (hence the *AR* prefix). Both tag systems are bi-tonal systems using markers made up of black and white square patterns, which seek to limit the effects of lighting variations on the tag detection process. ARToolKit encodes a feature vector of fixed length (usually 256 or 1024 elements) inside the black quadrilateral pattern outline. This vector is compared using correlation to a library of known markers and the presence of a tag can be detected by thresholding on the *confidence factor* output by the system (Owen [9] has modified the ARToolkit markers to include Fourier encodings). ARTag, on the other hand, applies digital techniques to encode and match patterns inside the square boundary of the tags and also uses *Forward Error Correction* (FEC) and *Cyclic Redundancy Check* (CRC) methods to encode 10 bits of information in a 36-bit binary sequence. The ARTag system relies on quadrilateral detection to identify the four corners of the tag boundary and then samples the image around the detected region using a 6×6 grid. This sampled interior is then checked using the CRC and FEC codes to retrieve the actual 10-bit binary sequence embedded within the tag. An additional bonus with both of these tag libraries is the ability to estimate orientation, which makes them suitable for camera calibration or pose estimation.

CyberCode [11] uses two-dimensional bar-codes as a basis for a marker system, placing particular emphasis on phase 3 of the operational chain described above. These can carry multi-bit labels, but their detection depends on their

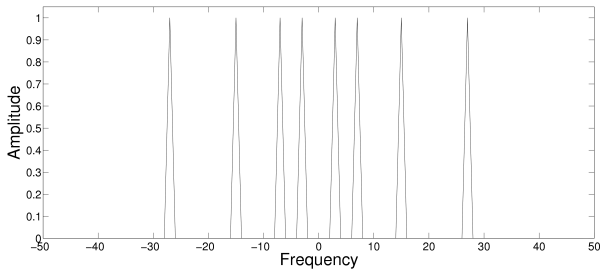


Figure 3. A binary representation of the number 210 in the frequency domain. The binary representation is 11010010.

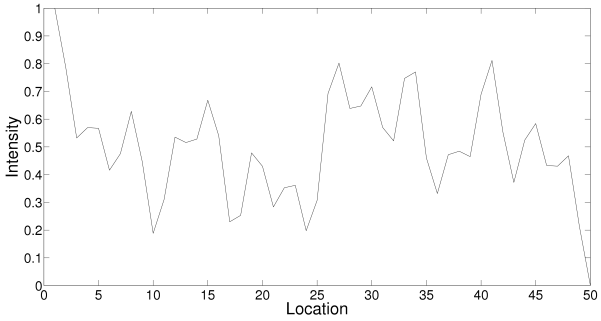


Figure 4. The signal in the spatial domain corresponding to the binary encoding of 210 shown in Fig. 3.

having good visibility in the image, however, if resolution is insufficient the data may not be intelligible even if the side bars that aid detection are found. Cybercode also has drawbacks as a fiducial marker system because it provides 3 salient points, and hence once the sidebars are located it can only correct for affine, not perspective, distortion.

One method that uses a multi-scale pattern is that of Cho and Neumann [1] whose targets are composed of a small set of colored concentric circles, with a slight resemblance to the markers developed in this paper. Their approach uses a fixed set of colored rings arranged in a square. Color is employed to ease detection but should also, in theory, allow for greater information encoding as compared to achromatic targets. On the other hand, printing costs, control of printer gamut mapping, and color constancy issues might make colored fiducials unattractive for many applications.

Claus and Fitzgibbon [2] developed a system of fiducials that employs machine learning to optimize the markers with respect to lighting conditions. This approach appears promising but is both computationally quite demanding, and also fails non-gracefully as resolution drops off.

3 Methodology

3.1 Fourier Tag Synthesis

The intent of our fiducials is to encode digital data suited to the end-user's particular application. In order to render the tags, we first encode the data by synthesizing a discrete periodic function using spectral coefficients $X(e^{j\omega})$. For simplicity, we represent our digital data as a binary encoding of evenly spaced, low-frequency signal *bursts*, slightly displaced from the DC component, as shown in Fig. 3. These coefficients correspond to the amplitude and phase of complex exponentials of harmonically-related frequencies present in the spatial domain according to the familiar formula-pair for the discrete-time Fourier transform:

$$x[n] = \frac{1}{2\pi} \int_{2\pi} X(e^{j\omega}) e^{j\omega} d\omega \tag{1}$$

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega n} \tag{2}$$

where $x[n]$ is a discrete aperiodic function, and $X(e^{j\omega})$ is periodic with length 2π. Equation 1 is referred to as the *synthesis* equation, and Eq. 2 is the *analysis* equation where $X(e^{j\omega})$ is often called the *spectrum* of $x[n]$.

For our application, we do not encode phase information, and therefore the $X(e^{j\omega})$ are purely real. This simplifies the decoding of the tag, since decoding phase information requires detecting the absolute phase of the signal, a task which is non-trivial. A side-effect of this decision is a less efficient encoding scheme, since we only use cosines and not both sines and cosines to encode information. This encoding scheme is a form of On/Off keying (OOK) modulation, where the information is stored in the presence or absence of a carrier signal. While it is not a particularly efficient modulation technique [6], its has the advantage of being fast and inexpensive to decode.

In order to map the corresponding signal in the spatial domain to pixel intensities, we must ensure that the signal in the spatial domain is also purely real. This important property can be established by constructing an even (symmetric) function in the frequency domain by mirroring $X(e^{j\omega})$, since an even, real signal in the frequency domain corresponds to an even, real signal in the spatial domain [8]. This signal, an example of which is shown in Fig. 4, can be readily obtained by using the fast Fourier transform (FFT).

In order to produce an arbitrary resolution digital image of the tag, we must first sample the resulting signal $x[n]$ to produce a *ray* of the tag, that is, a row of pixels emanating from the center of the tag. In practice, the signal is

Fourth Canadian Conference on Computer and Robot Vision(CRV'07)
0-7695-2786-8/07 $20.00 © 2007 IEEE

Figure 5. The Fourier tag of the number 210 **constructed from the signal shown in Fig. 4.**

scaled to be within the range of pixel values, and is filtered to reduce aliasing. The second stage of the process involves creating the full circular tag by rotating the ray about its origin. Due to the uniform discretization of pixels, some pixels from the rotated ray will potentially map to sub-pixel regions. We use sub-pixel stratified sampling with multiple samples in order to reduce this form of aliasing [10]. An example Fourier tag is shown in Fig. 5.

3.2 Detection

The first step in decoding a tag is to find its location in a given input image. Since the tags contain no color information, a gray scale image is sufficient for detection. Although the tags themselves are perfectly circular, perspective foreshortening will cause some degree of shape distortion, however, the shapes formed by the tags will always exhibit symmetry to some degree. Also, to decode the information encoded in the tags, we need to extract a ray from the tag that extends from the center to the tag boundary. Therefore, it is important to accurately find the center of the circular tag in the input image. Otherwise extraction of the ray would not be possible, thus complicating the decoding task.

The detection step is a two-stage process, the first of which is to find the center of the tag and the second is to extract a ray from the center to the periphery of the located tag. We first reduce the noise by blurring the input grayscale image and then threshold based on the average intensity in the image. This image is then passed on to a region detector which finds the center of all circular regions or *centroids*. The centroid locator finds the center of the circular regions in the image, including the actual tag center. This step can return false positives, requiring the second stage of the detection algorithm. To eliminate the majority of false positives that occur on uniformly illuminated regions of the image, we extract a strip of pixels along a line one pixel thick. The width of this line depends on the symmetry of the surrounding region. We grow the line in both directions starting from a potential tag center until pixels of both side differ significantly in their intensity values (a break in symmetry). The difference between the maximum and minimum intensities along the line are calculated in the range of pixel intensities. If the threshold is below a preset level, the point is discarded as a potential tag center. Running this process on the set of possible tag centers provides a set of candidate rays for decoding.

3.3 Decoding

Once a Fourier tag has been detected in an image, we need to extract the digital information encoded within it. This can be accomplished by first performing a Fourier transform on the detected ray described above. Since the relative displacement from the DC component as well as the relative spacing between the signal *bursts* of the binary encoding are known from the encoding stage, the rough location in the frequency domain for each bit can be determined. We can then inspect the frequency responses in the local neighborhood of each bit to determine whether the key is On or Off.

3.3.1 Sources of error

The performance of the decoding will be affected by many factors. Some are possible to minimize, while others are intrinsic to the setup and therefore cannot be minimized.

Imaging devices, either CCD or CMOS chips, inherently add pixel noise to the captured images. This noise is dependent on the cell itself, its temperature, as well as the equivalent ISO speed of the camera settings. It is often approximated as white noise.

Depending on the surrounding environment luminance, the tag itself might be under or over-exposed, effectively reducing the dynamic range of the tag image itself. This in turn increases the quantization noise. This noise is often approximated by white noise with variance proportional to the step size [12].

Frequency leakage happens when the number of cycles of the sampled signal fitting inside the sampling window is not an integer number. The latter will cause jumps at intervals equal to the size of the sampling window, under the assumption that the signal is periodic. The jumps arise from the fact that the end of the sampled signal does not connect smoothly with its beginning. In the spectrum, this

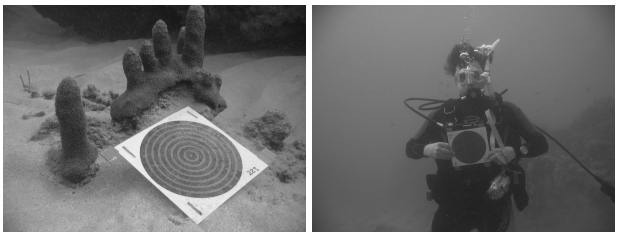


Figure 6. Fourier tags underwater on the sea bed and with a diver.

will show as smaller "echos" of the main components. This will affect the frequency resolution of the system, that is, the minimum distance for which two different frequencies can be resolved, and will ultimately influence the minimum distance between the frequencies used to store bits. This problem can be alleviated by multiplying the sampling window with a function that tapers at both ends. Commonly used windows include the Bartlett, the Hanning and the Parzen windows.

Underestimating the size of the tag would detrimentally affect the signal in a process similar to frequency leakage. This clipping of the tag blurs the spectrum and consequently reduces the frequency resolution of the system.

Overestimating the size of the tag introduces extra parts of the image inside the sampling window. If the intensity of this extra part is different than the average value of the tag, it will add low-frequency noise with roughly a $1/f$ distribution. This limits our ability to encode bits in the lower frequency area of the tag. This effect can be reduced by surrounding the tag with a grey value consistent with the average value of the tag, thus eliminating the abrupt change.

For all purposes, illumination changes inside indoor environments will only add low-frequency noise.

4 Human-Robot Interaction

While the focus of this paper deals with the design and encoding of the fiducial targets in the context of landmarks, they can also be used for interactive control. In our application, we employ fiducial markers to directly facilitate human-robot interaction. Specifically, we use the markers underwater to allow a scuba diver to communicate with a swimming robot vehicle to indicate desired actions or behaviors (see Fig. 6). This is especially important underwater since traditional human-robot interaction mechanisms such as speech and typing are infeasible, and radio communication is essentially impossible. Further, even traditional gesture-based interaction can be problematic since the cost of failures is high, feedback to verify the gestures is limited, and gestural vocabularies are typically small and error-prone. Using Fourier tags as a visual communication language combines robustness, expansive vocabulary, economy and simplicity. We have also considered the use of visual tags for vision-based control of terrestrial vehicles.

While the gestural language design we employ is outside the scope of this paper [4], it should be noted that the language is expressed in symbolic tokens represented by tags. The language structure is broken down into 3 classes of utterance:

1. Simple imperative commands: "stop", "surface now", "take a picture", etc.

2. Specification of numerical parameters: "double your current velocity", "increase controller gains by 10%", etc.

3. Extended programming scripts: "record a macro", "execute a macro", etc.

5 Experimental Results

In order to perform tests on real-life images within the laboratory, a number of Fourier tags were printed on letter-sized paper and mounted on the walls. Photographs were taken at various distances to replicate a robot moving in this environment and observing the tags.

The Fourier tag discussed in this section is the encoding of bit sequence 11010010, with the most significant bit encoded in the lowest frequency (f_1), and conversely the least significant bit encoded in the highest frequency (f_8). Figure 7 shows the result of using this particular Fourier tag at a relatively close distance (approx. 3m). The full image is seen in the top left corner, with the tag successfully detected. For the reader, the same tag has been reframed in the top right corner, with its location in the full scale image identified by the black box. A line traversing the tag through its center has been extracted and shown in the bottom left corner. The locations of bits 1 through 8 are shown on the amplitude spectrum in the bottom right corner. This particular graph shows that the bits can be easily decoded, with a signal to noise ratio of around 20 dB. The 1's are shown as peaks at f_1, f_2, f_4 and f_7 and the zeros at the other f_n locations.

As we move further away, graceful degradation of the tag is observed as the resolution of the tag decreases. This can be seen visually in the top right image of Fig. 8, where the tag's highest frequencies have been filtered out by the optics. With a net tag radius of 64 pixels in the image (only 0.07 % of the image area), frequencies f_6 through f_8 can no longer be detected (they are all encoded at frequencies above the Nyquist limit of 32). Frequencies f_1 through f_5 are still present and nicely detected, with a SNR ratio again of around 20 dB.

Fourth Canadian Conference on Computer and Robot Vision(CRV'07)
0-7695-2786-8/07 $20.00 © 2007 IEEE

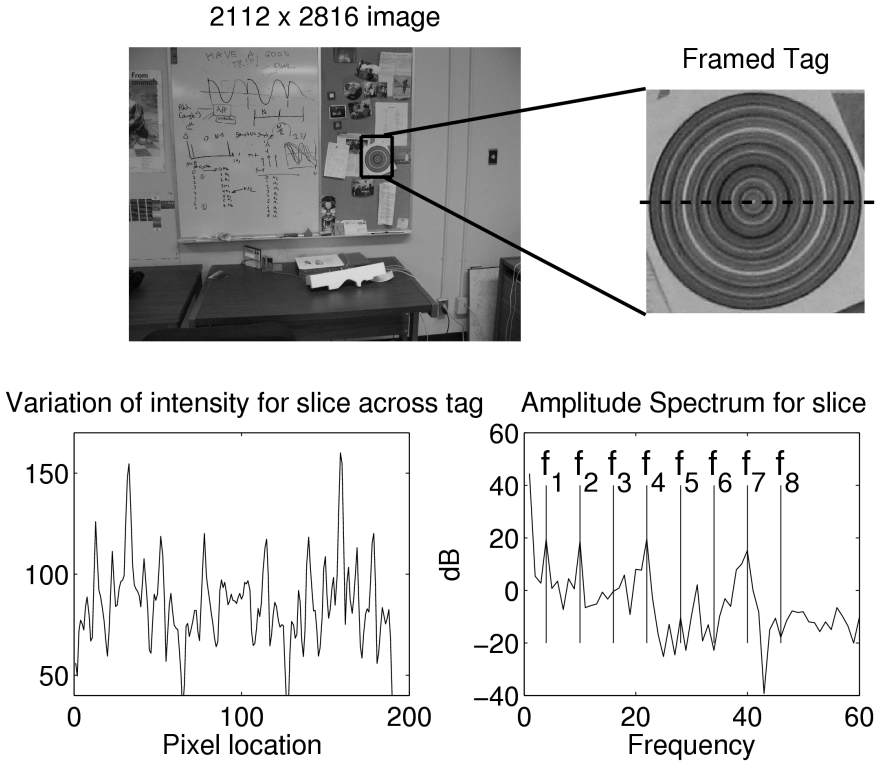


Figure 7. Picture of a Fourier tag at a distance of approx. 3 m. The right top image shows an enlarged section of the original picture. The bottom two graphs are pixel intensity across the dotted line, and its related Fourier transform. The vertical lines at f_1 through f_8 represent the location of the 8 bits.

6 Conclusions

This paper introduced a fiducial target system called the Fourier tag. This is a synthetic marker that can be used in robotics or virtual reality to provide readily detectable labels on objects. Critical issues for fiducial targets are detection, and for targets that carry an information payload, decoding of the encoded bits (i.e. the labels of the specific tag). Fourier tags are designed to be singularly easy to detect given the fact that they are completely rotationally symmetric. Information is encoded in the tags using a combination of circular harmonically-related functions with zero phase. The key property of the Fourier tag, in contrast to all prior fiducial marker systems, is that the payload carried by the tag degenerates gracefully with decreasing resolution. That is, as the camera-to-tag distance increases and the tag becomes less and less resolvable, the number of data bits that can be extracted gradually decreases. All comparable systems have a range beyond which the tag remains detectable, but for which the payload is no longer accessible. This degraded payload may be sufficient for some tasks. In addition, it can provide a robotic system sufficient data to determine if the tag should be approached for more complete information, or which of several tags should be more closely examined.

Our current tag detection technique does not attempt to handle occlusions; however, this would be possible by evaluating the symmetry from the tag center along several rays of varying angles. If searching in a given direction fails the symmetry test earlier than other directions, then an average of the rays of the longest length could be used. In addition, we have not used error correcting codes (ECC) in our current implementation, but would expect that we could increase the accuracy of decoding the payload data at the expense of encoding fewer bits.

One open research topic is the selection of schemes to encode a larger information payload in the Fourier tag. This could be accomplished by using mixtures of signals with different phases to encode the payload in the phase offsets. In principle this could yield substantially greater information density, but at the expense of both the simple appearance (and hence detectability) of the tag and the simplicity of the encoding and decoding process. Examining these tradeoffs remains a topic for further study.

Fourth Canadian Conference on Computer and Robot Vision(CRV'07)
0-7695-2786-8/07 $20.00 © 2007 IEEE

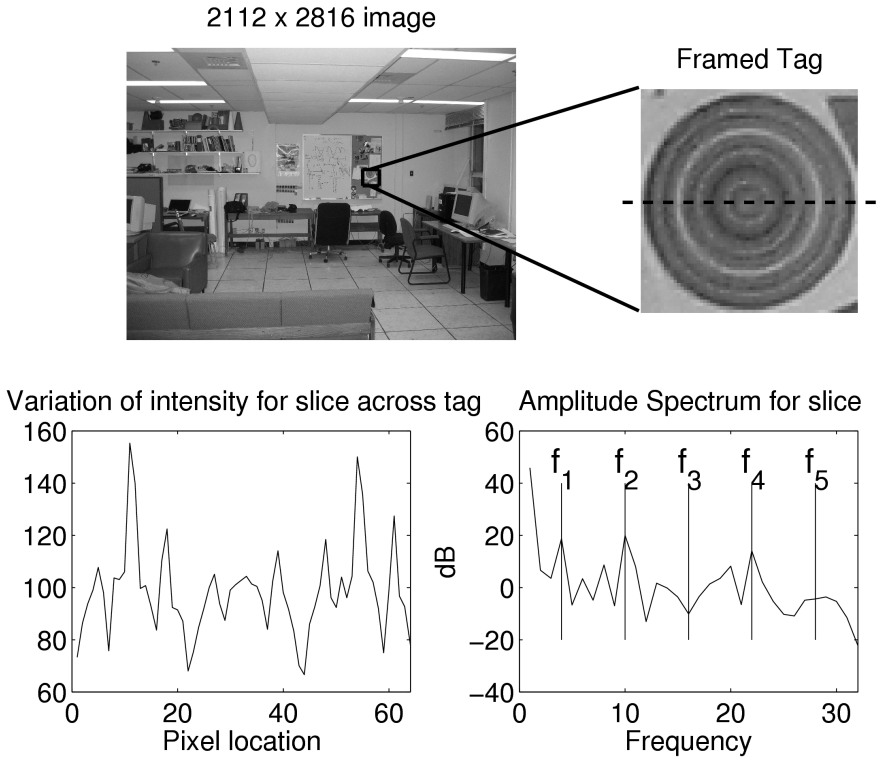


Figure 8. Picture of the same Fourier tag as in Fig. 7, but taken at a distance of approx. 6 m. The vertical lines at f_1 through f_5 again represent the location of the bits. Bits located at f_6 through f_8 have been lost.

It may also be possible to use color to increase the payload density as well as detectability, but this is complicated by the need to account for both illumination variations (and thus lack of color constancy), as well as printing variations that may make tag production more complicated.

References

[1] Y. Cho and U. Neumann. Multi-ring color fiducial systems for scalable fiducial tracking augmented reality, 1998.

[2] D. Claus and A. Fitzgibbon. Reliable fiducial detection in natural scenes, 2004.

[3] G. Dudek, P. Giguere, and J. Sattar. Sensor-based behavior control for an autonomous underwater vehicle. In *Proceedings of the International Symposium on Experimental Robotics, ISER*, Rio de Janeiro, Brasil, July 2006.

[4] G. Dudek, J. Sattar, and A. Xu. A visual language for robot control and programming: A human-interface study. In *Proceedings of the International Conference on Robotics and Automation ICRA*, Rome, Italy, April 2007.

[5] M. Fiala. Artag, a fiducial marker system using digital techniques. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 590–596, Washington, DC, USA, 2005. IEEE Computer Society.

[6] J. M. Geist. The cutoff rate for on-off keying. *IEEE Transactions on Communications*, 39(8):1179–1181, August 1991.

[7] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *In Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, San Francisco, October 1999.

[8] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals & systems (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[9] C. B. Owen, F. Xiao, and P. Middlin. What is the best fiducial? In *The First IEEE International Augmented Reality Toolkit Workshop*, pages 98–105, Darmstadt, Germany, Sept. 2002.

[10] M. Pharr and G. Humphreys. *Physically Based Rendering From Theory to Implementation*. Morgan Kaufmann, 2004.

[11] J. Rekimoto and Y. Ayatsuka. Cybercode: Designing augmented reality environments with visual tags, 2000.

[12] H. Taub and D. Schilling. *Principles of Communication Systems (4th ed.)*. McGraw Hill, 1986.

[13] G. Welch, G. Bishop, L. Vicci, S. Brumback, K. Keller, and D. Colucci. The HiBall tracker: High-performance wide-area tracking for virtual and augmented environments. pages 1–10.