

Mean-Subtracted Normalized Cross-Correlation Structured Light

Liviu-Mihai Calin Kiriakos N. Kutulakos
Department of Computer Science
University of Toronto
`{liviu,kyros}@cs.toronto.edu`

Abstract

We consider the code generation and code matching of a Structured Light 3D Scanning technique that uses only a few patterns. Phase Shifting codes have proven to be robust and accurate, while other types of codes require extra images for calibration, complex imaging systems, or a sacrifice in depth resolution to lower the number of images. We introduce Mean-Subtracted Normalized Cross-Correlation, a novel technique that is not related to Phase Shifting, but works with a limited number of images. The code matching is robust because albedo is naturally factored out by using normalized codes and effects of global illumination can be canceled out by subtracting the mean of captured images. The code creation assigns each pixel its own code and these codes are optimized to be as uncorrelated with each other as possible. The creation of codes is very flexible since the maximum frequency of the projected patterns can be limited and the number of projections used can be set. Due to the way the codes are created, the matching algorithm is just a mean, a subtraction, and a cross-correlation for each pixel. This leads to a very fast matching and can even be used to match codes from other techniques. We demonstrate that our technique works with at least 4 patterns using both simulated and real images.

1. Introduction

The creation of 3D models of real world objects is a popular topic that has uses in the entertainment, industrial, and medical industries. A widely applicable way of capturing the models is 3D scanning, which attempts to construct the model by viewing it without contact. The basic idea of 3D scanning is to capture 2D information from two or more views, find correspondences between the views, and compute the 3D information of those correspondences. The two views could be treated as cameras that use stereo vision to reconstruct 3D information. This method requires adequate calibration and complex point matching to find the correspondences. The much used alternative is to treat one of the

views as an inverse camera, or projector. Structured Light techniques project a series of patterns onto a scene, capturing the deformations of the patterns due to the depth of different objects. This allows for a known codification of each point in the scene, causing the correspondence matching to occur only between codes temporally, not between points spatially in the scene.

Many Structured Light techniques exist, but are created to work under ideal conditions using conventional projectors and cameras. New imaging systems [2, 13] can lift existing constraints on Structured Light techniques while introducing new ones. Interreflections within a scene can be made invariant or eliminated, and Structured Light patterns can be turned into one-shot patterns, but these may come at the expense of lower exposure, higher noise, and a restriction on the number of patterns. Even in conventional imaging systems, trading exposure time for noise and working in the presence of blur must be taken into account. This widens the space to search for optimal Structured Light patterns.

We present *Mean-Subtracted Normalized Cross-Correlation (MS-NCC)* as both a method to generate Structured Light patterns and a method to do correspondence matching between any Structured Light codes. MS-NCC is created with the objective of working under challenging imaging conditions. MS-NCC allows for the creation of codes of any length (at least 3, but demonstrated with 4 and above) that can allow for fast acquisition with short codes, or robust measurements with long codes. The created patterns that are projected can be banded in frequency to limit the effects of blur at the projector. The code matching algorithm uses only a mean, subtraction, and cross-correlation operation to find the correspondence of each pixel, so it is very fast. The matching algorithm is shown to work on any set of patterns, regardless if they were created with our technique or from another method.

Pribanić et al. [14] lists seven conditions that optimal Structured Light patterns and matching methods should satisfy. MS-NCC is near optimal in this respect.

1. The codes are contained within a projector pixel as

- they are temporal, not spatial, leading to a high resolution correspondence.
2. The goal of our code generation is to create codes that are as unique as possible from all other codes, including those of neighbouring pixels.
 3. Our code matching is robust to colour/albedo and constant global illumination at each pixel. As the frequency content of each pattern will contain both low and high frequencies instead of only a narrow band of high frequencies [11], diffuse interreflections may affect the result. It would not be as pronounced as having patterns with only low frequencies. While the codes themselves may not be robust to interreflections, we can eliminate the effects of interreflections at the hardware level.
 4. Each pixel in the resulting image can be matched independently so our code matching is robust to discontinuities.
 5. Our method will find the correspondences between projector and camera pixels, as many Structured Light techniques do, so determining the 3D coordinate between each correspondence can be done by any straightforward existing method.
 6. As mentioned above, the only processing done on the captured images, per pixel, is a mean of the captured code, a subtraction of that mean from the code, and a cross-correlation of the mean-subtracted code with the mean-subtracted normalized set of projected codes to find the correspondence as the code with the highest correlation.
 7. The contributions of this method is the creation of Structured Light patterns and the code matching of those patterns. This means MS-NCC can work in any Structured Light system. We show this by using an imaging system that can capture both conventional images and epipolar-only images [2].

One type of pattern that can meet the optimal requirements, while needing only a few number of patterns, is Phase Shifting. Each pattern is a single frequency cosine shifted a number of times. We wanted to expand the search space beyond a single frequency per pattern. To do so, an optimization is done on the whole space of patterns to create codes that are as unique as possible from all other codes. The only constraints are limiting the frequencies of patterns to not be forced to work at the Nyquist frequency and ensuring each code is normalized so that cross-correlation works properly.

The main contributions of the *Mean-Subtracted Normalized Cross-Correlation* method are:

- The use of MS-NCC for code matching, regardless of the coding scheme.
- The use of an optimization procedure for generating codes that are locally optimal according to the MS-NCC metric.
- A performance evaluation of different Structured Light coding schemes on both simulated data and real scenes. This comparison includes the current state-of-the-art schemes as well as MS-NCC.

2. Related

According to the classification of Structured Light techniques from [15], MS-NCC is most similar to a temporal n-ary code, but continuous. Despite the patterns being continuous, the matching does not solve for a relative or absolute phase. Instead, each pixel is assigned a code and its match is found by taking the most similar code from the set of codes.

A discrete, temporal n-ary code that uses a small number of patterns is the Hilbert code [7]. It is similar to the Binary code where each projector column is given an index, but it uses gray levels to reduce the number of patterns needed. The codes are created by finding the grid points of a low-dimensional low-order Hilbert curve. If the resolution of the patterns are not high enough, extra codes are created by interpolating between existing codes. While this allows for control of the number of patterns used, their frequency content, and the uniqueness of the initially selected codes, the interpolated codes, which attempt to give high resolution, are not optimal. Also, at least two extra projections are needed to determine the white and black level of the scene.

Efficient Multiple Phase Shifting [14] provides a criteria for optimal Structured Light techniques and creates multiple phase shifting codes whose frequencies are chosen to make phase unwrapping simple and robust. 8 images per frequency are needed, plus 2 images for reference.

Micro Phase Shifting (MPS) [5] and Embedded Phase Shifting (EMB) [10] both attempt to lower the number of patterns needed. Both methods use cosines in a narrow range of frequencies to be invariant to amplitude variation from projector defocus and consist of high frequencies to ignore the effects of diffuse interreflections. This can only handle a very low amount of blur, since high frequencies are the first to be attenuated by a significant amount of blur. We want to create patterns that do not need to have high frequencies, so they will be less affected by blur. MPS chooses its set of frequencies from optimizing the weighted difference between codes. EMB embeds the unit frequency and an exponentially increasing set of codes in a set of high frequency codes. Both methods start with a cosine shifted

three times. MPS adds an unshifted cosine for every additional frequency, while EMB adds two or three shifted cosines for every additional frequency. MPS uses a lookup table to find a camera pixel's correspondence in projector pixels, then uses the relative phase from the first frequency to find a subpixel correspondence. EMB can find a relative phase for each frequency and does pairwise phase unwrapping to get a set of correspondences for each camera pixel. Both of these methods have a low minimum number of patterns - MPS works with at least 4 patterns, while EMB with at least 5.

Elimination of global illumination can be done at the hardware level by applying Homogeneous Codes to both the projector and camera [12]. Although requiring a special imaging system, Structured Light codes can be projected and only the epipolar component of light captured. We use the Episcan3D [2] imaging system to test the effects of global illumination on the code matching of various Structured Light methods.

3. Image Formation Model

Let us consider corresponding pixels p on the projector and c on the camera. Let \mathbf{f}_p be the vector containing the code at projector pixel p . The length of \mathbf{f}_p is determined by the number of projected patterns used. Assuming that the scene is textureless, the strength of the signal from the projector is preserved at the camera, and there is no global illumination, the resulting image vector at camera pixel c is:

$$\mathbf{I}_c = \mathbf{f}_p. \quad (1)$$

This can be seen in Figure 1. This assumes that no more than one projector pixel contributes to one camera pixel. In reality, it is very tough to have a one-to-one mapping between projector and camera pixels.

If we now consider texture or albedo in the scene and the intensity difference between the projector and camera, this can be treated as a scalar factor at each pixel, a_c . There will also be an additive offset, o_c , that is the contribution of global illumination, such as ambient light. These values will be treated as constant for each projection at that pixel, but may vary across pixels in the image. In reality, global illumination is not constant, but we will be using a hardware-level solution to make global illumination largely independent of the projected patterns. The resulting image vector at camera pixel c is now:

$$\mathbf{I}_c = a_c * \mathbf{f}_p + o_c. \quad (2)$$

3.1. Noise Model

There are various sources of noise that will affect the captured image [6]. These are photon shot noise, read noise, saturation noise, and quantization noise.

Photon shot noise or Poisson noise arises from the particle nature of light. Photons will hit the camera sensor causing photoelectrons to be emitted by the photoelectric effect. The camera sensor collects and counts these electrons to assign an intensity value in the image. Since the arrival of photons is random, the number of electrons counted by the sensor may vary over a given time period. This variation can be modeled by a Poisson distribution. The image vector at camera pixel c is a Poisson random variable whose mean is the expected vector from Equation 2 in electrons:

$$\mathbf{I}_c = \text{Poisson}(k_1 * (a_c * \mathbf{f}_p + o_c)), \quad (3)$$

where k_1 is the conversion factor of electrons per intensity and each element in the vector is its own Poisson random variable.

Read noise comes from the fact that the light signal cannot be measured perfectly. The electronics of the camera sensor will introduce a spurious signal. This is modeled by a Gaussian distribution with a zero mean and a camera-dependent standard deviation. Because of this, read noise is most noticeable in low light conditions. The image vector at camera pixel c after accounting for read noise is:

$$\mathbf{I}_c = \text{Poisson}(k_1 * (a_c * \mathbf{f}_p + o_c)) + \mathcal{N}(\mathbf{0}, \sigma_r^2), \quad (4)$$

where σ_r is a vector containing the read noise standard deviation in electrons to give a different result for each image.

Saturation noise acts as a clamping due to the sensor reaching an upper limit on the signal it can record. Each pixel will store photoelectrons due to incoming photons, but it cannot store more electrons than its full well capacity. The image vector at camera pixel c after accounting for saturation noise is:

$$\mathbf{I}_c = \min(\text{Poisson}(k_1 * (a_c * \mathbf{f}_p + o_c)) + \mathcal{N}(\mathbf{0}, \sigma_r^2), k_2), \quad (5)$$

where k_2 is the full well capacity in electrons.

Quantization noise acts as a rounding as the image intensity is generally represented as an 8-bit to 16-bit value, not a very large number of electrons. The image vector at camera pixel c after accounting for quantization noise is:

$$\mathbf{I}_c = \lfloor k_3 * \min(\text{Poisson}(k_1 * (a_c * \mathbf{f}_p + o_c)) + \mathcal{N}(\mathbf{0}, \sigma_r^2), k_2) \rfloor, \quad (6)$$

where k_3 is the conversion factor of output intensity per electron.

4. Code Matching

The basis of Structured Light 3D scanning is to find correspondences between 2D views of the scene and then intersect the rays extending from the centers of projection to the corresponding pixels to find a 3D point. A correspondence

occurs if a camera pixel is illuminated directly by a projector pixel. Influences from other projector pixels through interreflections or blur should not affect the correspondence. Correspondences are found by projecting multiple patterns so that each pixel is coded differently. The effects of noise will be ignored.

Consider the simple image formation model from Equation 1. One way of attempting to match a camera pixel and a projector pixel is to correlate the image vector at a camera pixel with a code at a projector pixel. Applying the correlation between the image vector, \mathbf{I}_c , and a general projector code, \mathbf{f}_j , where j is any projector pixel is:

$$\langle \mathbf{I}_c, \mathbf{f}_j \rangle \quad (7)$$

$$= \langle \mathbf{f}_p, \mathbf{f}_j \rangle. \quad (8)$$

If all codes are normalized, the correlation is 1 when $j = p$, otherwise the correlation will be less than 1.

Consider the image formation model from Equation 2 that does not contain noise, but accounts for a constant multiplicative factor and a constant offset. Applying the correlation directly between the image vector, \mathbf{I}_c , and a general projector code, \mathbf{f}_j , where j is any projector pixel is:

$$\langle \mathbf{I}_c, \mathbf{f}_j \rangle \quad (9)$$

$$= \langle a_c * \mathbf{f}_p + o_c, \mathbf{f}_j \rangle \quad (10)$$

$$= \langle a_c * \mathbf{f}_p, \mathbf{f}_j \rangle + \langle o_c, \mathbf{f}_j \rangle \quad (11)$$

$$= a_c * \langle \mathbf{f}_p, \mathbf{f}_j \rangle + \langle o_c, \mathbf{f}_j \rangle. \quad (12)$$

The additive constant is not factored out, which may cause the highest correlation to incorrectly occur when $j \neq p$.

Instead, it can be eliminated from the correlation by subtracting out the mean of the image vector:

$$\mathbf{I}_c - \text{mean}(\mathbf{I}_c) \quad (13)$$

$$= a_c * \mathbf{f}_p + o_c - (a_c * \text{mean}(\mathbf{f}_p) + o_c) \quad (14)$$

$$= a_c * \mathbf{f}_p - a_c * \text{mean}(\mathbf{f}_p) \quad (15)$$

$$= a_c * (\mathbf{f}_p - \text{mean}(\mathbf{f}_p)). \quad (16)$$

If we now correlate between the mean-subtracted image vector and a mean-subtracted general projector code, the result is the *Mean-Subtracted Normalized Cross-Correlation (MS-NCC)*:

$$\langle \mathbf{I}_c - \text{mean}(\mathbf{I}_c), \mathbf{f}_j - \text{mean}(\mathbf{f}_j) \rangle \quad (17)$$

$$= \langle a_c * (\mathbf{f}_p - \text{mean}(\mathbf{f}_p)), \mathbf{f}_j - \text{mean}(\mathbf{f}_j) \rangle \quad (18)$$

$$= a_c * \langle \mathbf{f}_p - \text{mean}(\mathbf{f}_p), \mathbf{f}_j - \text{mean}(\mathbf{f}_j) \rangle. \quad (19)$$

If the mean-subtracted codes are normalized, the correlation is as before, giving the highest value when $j = p$. The constant scalar factor will scale each correlation by the same amount, so it will not change the correlation order.

Both a constant multiplicative scalar and additive offset can be handled by a mean-subtracted normalized cross-correlation. If either of those values are not constant, they are not factored out and will cause problems with the matching.

This gives rise to our matching algorithm. Let

$$\Pi = [\mathbf{f}_1 \ \mathbf{f}_2 \ \mathbf{f}_3 \ \cdots \ \mathbf{f}_N], \quad (20)$$

be the set of Structured Light codes represented as a matrix. N is the total number of Structured Light codes, one for each projector pixel. Figure 2 shows how Π is used in the image acquisition. First, prepare each Structured Light code for correlation by subtracting out its mean and normalizing:

$$\mathbf{Q} = [\mathbf{f}_1 - \text{mean}(\mathbf{f}_1) \ \mathbf{f}_2 - \text{mean}(\mathbf{f}_2) \ \cdots \ \mathbf{f}_N - \text{mean}(\mathbf{f}_N)], \quad (21)$$

$$\mathbf{Q}_{\text{norm}} = \left[\frac{\mathbf{f}_1 - \text{mean}(\mathbf{f}_1)}{\|\mathbf{f}_1 - \text{mean}(\mathbf{f}_1)\|} \ \cdots \ \frac{\mathbf{f}_N - \text{mean}(\mathbf{f}_N)}{\|\mathbf{f}_N - \text{mean}(\mathbf{f}_N)\|} \right]. \quad (22)$$

Then, for each pixel in the set of images and its image vector \mathbf{I}_c , subtract out its mean and correlate the mean-subtracted image vector with the mean-subtracted normalized Structured Light code matrix. The index of the highest correlation corresponds to the projector pixel with the best match:

$$\arg \max_j ((\mathbf{I}_c - \text{mean}(\mathbf{I}_c))^T \mathbf{Q}_{\text{norm}}), \quad (23)$$

where j is the column index of the resulting row vector.

The precision of the matching depends on the resolution of the projector since the matches can only happen with a pixel, not a combination of pixels. Phase Shifting techniques assign continuous phases to each camera pixel, allowing for sub-pixel matching with the projector.

5. Code Generation

The main goal of MS-NCC's code generation algorithm is to label each pixel with a code that is as unique as possible, where the uniqueness metric used is mean-subtracted normalized cross-correlation, as explained in Section 4. We first explain the concept under normalized cross-correlation (NCC), to show the algorithm without the added complexity of mean subtraction.

We want the correlation of a given code from one pixel with a code from any other pixel to be as low as possible, regardless of how far the pixels themselves are. As there is less confusion between codes, the number of incorrectly matched pixels should be less. Quantitatively, we want the maximum value of every code correlated with every other code to be minimized, apart from itself, which is always 1 when using normalized codes. This amounts to minimizing the infinity norm of the set of inner products between all pairs of codes.

Note that the infinity norm in the context of this algorithm is an entry-wise norm. This is equivalent to first vectorizing a matrix and then applying the vector infinity norm:

$$\|\boldsymbol{\Pi}^T \boldsymbol{\Pi} - \mathbf{I}\|_\infty \quad (24)$$

$$= \|vec(\boldsymbol{\Pi}^T \boldsymbol{\Pi} - \mathbf{I})\|_\infty \quad (25)$$

$$= \max_{ij} |(\boldsymbol{\Pi}^T \boldsymbol{\Pi} - \mathbf{I})_{ij}|, \quad (26)$$

where the $M \times N$ projection matrix, $\boldsymbol{\Pi}$, is as defined in Equation 20.

The $N \times N$ correlation matrix $\boldsymbol{\Pi}^T \boldsymbol{\Pi}$ will contain the set of inner products between all pairs of codes. If $\boldsymbol{\Pi}$ is normalized, the main diagonal of the correlation matrix will all be ones. Ideally, the off-diagonal elements will be small to avoid confusions between them in the presence of noise. We want the correlation matrix to be as close to the Identity matrix as possible, but are limited by its rank, since $M < N$.

We also want to be able to limit the frequency of the projected patterns. This will allow for projection of lower frequencies to counteract the effects of blur. The values along the rows of $\boldsymbol{\Pi}$ must be frequency limited, as these rows will be projected onto the scene.

This leads to the optimization:

$$\min \|\boldsymbol{\Pi}^T \boldsymbol{\Pi} - \mathbf{I}\|_\infty \quad (27)$$

$$\text{s.t. } \boldsymbol{\Pi} \geq 0, \forall j \|\mathbf{f}_j\|_2 = 1, \mathbf{D}_{\text{high}} \boldsymbol{\Pi}^T = 0. \quad (28)$$

$\boldsymbol{\Pi} \geq 0$ must hold since negative light cannot be projected. $\|\mathbf{f}_j\|_2 = 1$ makes sure every code is normalized so that NCC works properly. $\mathbf{D}_{\text{high}} \boldsymbol{\Pi}^T = 0$ enforces the frequency constraint, where \mathbf{D}_{high} is the Discrete Fourier Transform matrix that only returns the values for frequencies higher than the maximum frequency threshold.

In practice, what matters is not how unique the projected codes are, but how unique the codes will be when observed at a camera pixel. These codes will be affected by albedo and global illumination.

Albedo is naturally taken care of by NCC, since a multiplication by a constant scalar will factor out of the inner product. The addition of a constant scalar does not factor out and will lead to the wrong correlation result. Section 4 shows one strategy for removing the additive offset by subtracting out the mean of each code. As long as the offset is constant for each value in one code, subtracting the mean will eliminate the offset. This means we must correlate not normalized codes, but mean-subtracted normalized codes. The new correlation matrix is $\mathbf{Q}^T \mathbf{Q}$ where \mathbf{Q} is defined in Equation 21.

Thus, the new optimization is:

$$\min \|\mathbf{Q}^T \mathbf{Q} - \mathbf{I}\|_\infty \quad (29)$$

$$\text{s.t. } \boldsymbol{\Pi} \geq 0, \forall j \|\mathbf{f}_j - \text{mean}(\mathbf{f}_j)\|_2 = 1, \mathbf{D}_{\text{high}} \boldsymbol{\Pi}^T = 0. \quad (30)$$

$\boldsymbol{\Pi} \geq 0$ and $\mathbf{D}_{\text{high}} \boldsymbol{\Pi}^T = 0$ are the same constraints since the mean subtraction happens only after the images are captured, not before the projection happens. $\|\mathbf{f}_j - \text{mean}(\mathbf{f}_j)\|_2 = 1$ makes sure that the mean-subtracted codes are normalized, since the correlation is done on them.

5.1. Optimization Details

The optimization is done with Python’s Sequential Least Squares Programming (SLSQP) optimizer found in `scipy.optimize.minimize` using a random initial guess that satisfies all constraints. The optimization was not done using the infinity norm, but done using the 100-norm, defined as:

$$\|\mathbf{A}\|_{100} = \left(\sum_{i=1}^M \sum_{j=1}^N |a_{ij}|^{100} \right)^{1/100}. \quad (31)$$

This approximates the infinity norm, but runs about 17 times faster. This lets an optimization of $N = 512$ pixels run on the order of days instead of months.

5.2. Optimized MS-NCC Codes

An example of the optimized MS-NCC codes are shown in Figure 3. Contrast that with the types of patterns and correlation matrices that occur when using Phase Shifting methods, as seen in Figures 4, 5, and 6. By not being limited to single frequency sinusoidal patterns, the space of available codes is expanded and unstructured codes can be used while still providing a measure of uniqueness.

We optimized our set of codes for 512 codes, code lengths of 4, 5, 6, and 7, and for each maximum allowable integer frequencies between 1 and 64. The infinity norm of the resulting $\mathbf{Q}_{\text{norm}}^T \mathbf{Q}_{\text{norm}} - \mathbf{I}$ for each case can be seen in Figure 7. Note that for our codes, \mathbf{Q} and \mathbf{Q}_{norm} are equivalent because of the normalization constraint in the optimization. \mathbf{Q}_{norm} is explicitly used to be able to compare against other methods’ codes using the same metric. Also shown are the infinity norms for the Phase Shifting codes described in Table 1.

Patterns with infinity norms comparable to the state-of-the-art codes have been created when starting from a random guess (within the constraints of the optimization) in real space. We achieve the best infinity norm using 6 patterns and high frequencies, and 7 patterns and mid to high frequencies.

Figure 9 shows which pattern has the lowest infinity norm for all frequencies up to the maximum allowable frequency. Although MS-NCC codes do not have the lowest infinity norms in all cases, we can use the lower infinity norms envelope of patterns to test if the infinity norm metric will lead to the lowest errors.

Two main problems with our optimization are that the search space is huge, the number of pixels multiplied by the

number of patterns, and that the constraints on the optimization can be harsh. Single and Multiple Phase Shifting codes only need one frequency to create one pattern unlike ours that optimizes the patterns in real space. The constraints of normalizing each code and eliminating high frequencies may cause steps taken by the optimizer to be in an invalid state. Projection back onto a valid state will weaken those steps. In these cases, our optimized codes may not yield a lower infinity norm than other codes.

5.3. Optimization In The Frequency Domain

Rather than working in the spatial domain where the solution space is MN in size, we attempted to work in the frequency domain where the solution space is $M(2k + 1)$, where k is the maximum allowable frequency, $2k$ is from the complex-valued Fourier coefficient for each frequency, 1 is the DC component, and M for the number of patterns optimized. When the optimization converges to a local minimum, it is able to get a similar result as working in the spatial domain, but in a fraction of the time. The downside is that the optimization terminates on the first step if the maximum allowable frequency is less than $N/4$. As our interest in the range of frequency is $N/512$ to $N/8$, optimizing in the frequency domain is not an option. See Table 2 for test results of optimizing in both domains.

5.4. Adjusting The Initial Guess

The initial guess of the optimization is important here because we are not minimizing a convex function. The final result depends on where it starts. Also, there exist viable Structured Light codes already, such as from Phase Shifting techniques, so it may be possible to start from those and discover a more optimized set of codes.

In all cases, when starting from a Phase Shifting code, the optimization ran for a few steps at most. The initial guess was at or near a local minimum. For a small number of pixels, the infinity norm was better for the optimized patterns starting with Phase Shifting codes than with a random initial guess. As we increased the number of pixels and the maximum allowable frequency in our pattern, the local minima around Phase Shifting codes was not as low as optimizing from a random initial guess.

Starting from an MPS code had a better optimization for a small number of pixels, but lost its advantage at a faster rate than starting from a general Phase Shifting code. Starting from an MPS code also had no speed up compared to starting from a random initial guess.

The optimization results for starting at these initial guesses can be seen in Table 3.

The optimization can be improved by starting from a previously optimized solution. This is more than just increasing the number of iterations. The new initial guess is set to the result that had the best infinity norm out of all of

the previous MS-NCC results, excluding those whose maximum allowable frequency is too high. The improvement to the infinity norm can be seen in Figures 7 and 8.

5.5. Summary Of Code Generation And Optimization

Despite our best efforts we were not able to go below the infinity norms of state-of-the-art codes for small code lengths and low frequencies. This suggests that there is still room for improvement in the optimization.

No single coding scheme produced consistently low infinity norm measurements. Codes specifically optimized with the infinity norm may have to switch between different existing coding schemes to attain the best possible infinity norm, as shown in see Figure 9.

6. Simulations

To compare the effectiveness of different coding schemes, we simulated a very simple scene - a textureless fronto-parallel plane. For each $M \times 512$ Π matrix, its rows were repeated to form $M 512 \times 512$ 2D projection patters. This was done for each set of codes, code lengths of 4, 5, 6, and 7, and maximum allowable frequencies of 1 to 64 (periods of 512 to 8 pixels, respectively).

To properly assess the expectation of errors, tens or hundreds of trials would need to be run. Due to the volume of data simulated and the processing time, this was not feasible. For each case, only one trial was run. The resulting error plots are not as smooth and clear as they could be.

The creation of images at the camera involved:

- Quantizing the signal at the projector.
- Adding a constant 10% of the maximum strength of the projector.
- Convolving a 2D circle of diameter 8 pixels to blur the pattern, or skipping the convolution to not blur the pattern.
- Warping the pattern according to the projector-to-image homography associated with that scene plane.
- Computing the intensity at an image pixel by integrating the warped pattern over the pixel's footprint.
- Scaling the signal based on the level of exposure of the camera.
- Adding noise.
- Quantizing the signal at image.

The noise was determined from a Canon 1D Mark II camera model [1]. At ISO 100, the maximum signal at a

pixel in electrons is 53000. Assuming the simulated camera’s exposure was set such that a pixel’s maximum exposure corresponded to 53000 electrons, the value chosen at a camera pixel is taken from a Poisson distribution with a mean of the expected values scaled by 53000 over the maximum exposure. The read noise for this camera model is 16.61 electrons at ISO 100. Thus, the additive read noise was modeled by a Gaussian distribution of mean 0 and standard deviation 16.61.

Nine levels of exposure were simulated. One was set so that the camera and projector had the same exposure, barring changes from blurring and noise. Four were set by decreasing the exposure by one stop each ($\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, and $\frac{1}{16}$ of the maximum exposure). Four were set by increasing the exposure by 1.25, 1.5, 1.75, and 2 times the maximum exposure. In the case of underexposure, quantization at the camera plays a bigger role, while for overexposure, saturation is the main source of noise.

Figure 10 shows an example of the scene as imaged under one projection pattern for a variety of imaging conditions.

6.1. Evaluation And Results

For each type of pattern we compared against, its respective matching method was used. For the Phase Shifting patterns, a continuous matching can be achieved, whereas our method can only give a discrete matching. MS-NCC matching was also applied to each compared pattern. This gave comparable results, even for codes not optimized in this way or codes that are not naturally normalized.

The Middlebury Stereo Evaluation [16] was used to evaluate the matched disparities. The two metrics applied are the percentage of incorrect pixels (with a 0.5 pixel threshold), and the root mean squared (RMS) disparity error. An incorrect pixel is defined as a pixel whose absolute disparity error is greater than 0.5. The number of incorrect pixels to total pixels, written as a percentage, is the percentage of incorrect pixels.

Figure 11 shows the percentage of incorrect pixels for the ideal exposure. Due to the low level of noise, there are not many errors except for the patterns using very low frequencies. For those, it is very easy to confuse a code with the one beside it. The corresponding RMS can be seen in Figure 13.

In the case of underexposure, the trends of the infinity norms from Figure 7 can be seen in the percentage of incorrect pixels in Figure 15. MS-NCC codes have low errors for very high frequencies for 5 patterns, the lowest errors for high frequencies for 6 patterns, and low errors and the lowest above the middle frequency for 7 patterns. EMB and COSN, which have the lowest infinity norms for low frequencies for 5 and 6, and 6 and 7 patterns, respectively, also have the lowest errors at the same frequencies. COSU that

relies solely on the unit frequency has very high errors. The main exception to the trend is MPS. MPS has a poor infinity norm but has a low number of incorrect pixel for all pattern sizes. Figure 17 shows how patterns with few incorrect pixels also have low RMS. The exceptions are patterns of very low frequency, especially the unit frequency, that have lots of incorrect pixels because the codes are similar, but a lower RMS because of good localization properties. Figures 19 and 20 show the effect of a circular blur of an 8 pixel diameter on both error metrics. As higher frequencies are blurred out, the error rises greatly, while those close to the unit frequency are unchanged. Blur is especially detrimental to MS-NCC. For low frequencies, MS-NCC has a higher infinity norm and higher errors than the other patterns, while its infinity norm advantage is eliminated for higher frequencies.

Figures 21 and 22 show the errors in the presence of saturation. MS-NCC clearly has a lower occurrence of errors, but that does not translate into a lower RMS as it is not optimized for that.

Under ideal conditions and 6 or 7 patterns, any non-unit frequency is good throughout. For 5 patterns, MPS and EMB have the lowest incorrect pixels and RMS errors. For 4 patterns, COS1 has good RMS for high frequencies. Despite the high percentage of incorrect pixels, MPS that uses sub-unit frequencies has the best RMS for 4 patterns.

With underexposure and 4 or 5 patterns, the best patterns are the same as above. For 6 patterns, EMB and COSN are best for low frequencies, COS1 is good for mid-low frequencies, and MS-NCC is good for high frequencies. For 7 patterns, EMB is best for the lower half of frequencies, while MS-NCC is best for the higher half of frequencies.

The 8 pixel circular blur will leave low frequencies almost unchanged, while high frequencies are affected greatly. In the presence of blur, the unit frequency patterns will start to have a better RMS. If the blur is increased, they will have a much better RMS while having a comparable incorrect number of pixels, while the higher frequency patterns have more errors.

For overexposure, MS-NCC has the lowest number of incorrect pixels, while COSU has the lowest RMS.

When comparing code matching methods, the lack of matrix inversion and phase unwrapping makes MS-NCC at least 5 times faster than all the other methods.

We have seen in practice that our MS-NCC matching applies to any set of Structured Light codes and gives similar results to that Structured Light codes’ specialized matching algorithm. Even methods whose mean-subtracted codes are not normalized or optimized to be unique, as most are not, work with MS-NCC as long as each mean-subtracted code is explicitly normalized.

7. Experiments

We want to capture images with a real imaging system to test whether the simulation results are true in practice. The imaging system used can minimize the effect of indirect illumination, thus putting all patterns at a roughly equal footing as far as interreflections are concerned.

The Episcan3D system [2] is used to capture images for our experiments. The projector used is the MicroVision ShowWX+ Laser Pico Projector, as shown in Figure 23. Its usable resolution is 800×400 , but only the first 512 pixels were used as that is the length of our patterns. Each of the patterns were repeated 400 times along the rows to create a 2D projection. The camera used is an IDS USB 3 uEye CP camera with a resolution of 1600×1200 . Along with conventional images, this system can capture epipolar-only images at the hardware level. Rather than a pixel capturing all of the light in the scene, it will capture only the light along its epipolar plane. This means an epipolar image will mainly capture direct reflections, not interreflections and subsurface scattering.

The three objects used are a flat board, a flat wedge that has strong interreflections around the edge, and a bowl that has weaker, but more prevalent, interreflections. The flat board is chosen because it represents the most ideal object. The amount of global illumination should be very low, the object structure is very simple, and it is the same shape as the iso-depth plane used in simulations so the results should be connected. The flat board will not contain any diffuse interreflections. The diffuse flat wedge and the diffuse bowl are used to capture the effects of diffuse interreflections. The sharp wedge has stronger but localized interreflections, while the round bowl has weaker but widespread interreflections. Using the capabilities of the imaging system, the diffuse interreflections can be largely eliminated without relying on the patterns themselves. This will allow low frequency patterns to image the traditionally troublesome objects.

Objects with specular reflections or subsurface scattering, and objects with complex shapes are not imaged. We did not include more complex geometries because we had no means to obtain ground truth shape information to compare against.

7.1. Ground Truth

To be able to quantitatively compare results from various Structured Light methods, the ground truths of the objects imaged are needed. This is generally done by projecting a state-of-the-art robust pattern, such as Gray codes [8], from the imaging system used, or by imaging an object with a known 3D model. Using an object that matches its 3D model requires high accuracy equipment to manufacture or a high cost to acquire. There are also tools to capture high accuracy scans of real-world objects, but these

requires extra resources [9]. Having a 3D model as ground truth means there must be a step for transitioning between image space and 3D space, or visa versa. This requires an accurate calibration of the imaging system.

These extra steps can be avoided if the ground truth can be measured by a Structured Light pattern. Comparing to the ground truth only involves comparing the stereo disparities. There are downsides to using a Structured Light pattern for ground truth. High accuracy patterns generally require high frequency patterns, which means there should not be blur in the imaging system. The images must be well exposed to be robust to noise. Objects with specular reflections and subsurface scattering will not be handled well. Even with an imaging system that can capture epipolar-only images, if there is any non-epipolar light let in, this will affect the result.

These sorts of imperfections are present in our imaging system: a dim image with the highest power from the projector and largest aperture at the camera, laser speckle causing a false texture and more dimming in some pixels, a blurred laser projector, and errors in the epipolar geometry at the hardware level causing a small field of view for epipolar-only imaging, non-epipolar leakage, and jitter causing intensity variation.

Figure 24, Figure 25, and Table 4 show a high amount of jitter present in the system. This is a bigger problem for epipolar-only imaging rather than conventional imaging since the geometric constraints that allow the system to do epipolar-only imaging will not always hold. This leads to a high amount of errors in both imaging modes, with even higher errors for epipolar-only in cases where the opposite should hold.

In this imaging system, using a large set of high frequency patterns, such as Gray Codes, for the ground truth will suffer due to blur. Using a low frequency alternative, such as a Multiple Phase Shifting code with 16 low frequencies (1 to 16) shifted 10 times each, will also have problems from the non-epipolar leakage. Both sets of codes were captured and although there was disagreement between the two, it was only on the order of a pixel. Since the objects imaged were simple, continuous objects, it was possible to fit a plane to the board, two planes to the wedge, and a plane and sphere to the bowl. Using Iteratively Reweighted Least Squares [4] to robustly fit the above shapes to the objects completely agreed between the two Structured Light methods. Ultimately, the results in the following section must be taken with a grain of salt.

Acquiring ground truths for objects of more complex shapes or light paths would not be possible with Structured Light methods in the current imaging system. Using a known 3D model would also not work in this system. The calibration that will allow for transformation between 3D space and image space still requires patterns to be projected

and images to be captured. The best reprojection error after calibrating multiple times was greater than half a pixel. This means that perfect pixel correspondences will have significant error in 3D space, or the 3D model projected into image space will not cover the correct pixels.

These problems greatly limited the possible objects we could image.

7.2. Evaluation And Results

A subset of frequencies were used due to the amount of time and resources it took to work with real data. Only one exposure was taken due to hardware limitations.

An example of the matching for each set of codes and each target can be seen in Figures 26, 27, and 28. The high number of errors seen are due to imperfections in the imaging system.

Figures 29 and 30 compare errors from using the MS-NCC matching method and each pattern's specialized method on images of the flat board. Again, it is shown that MS-NCC can be used on any pattern regardless if it is optimized for it or not. The effects of blur can be seen on high frequencies, while the low frequencies are relatively less affected and have better errors. Notice how MS-NCC no longer gives the lowest errors for a high pattern number and a high frequency due to the blurring. Although the conditions of the imaging system may not exactly match the simulated conditions, the errors from imaging a flat board are similar to the errors from the simulated, underexposed, and blurred iso-depth plane. Conventional and epipolar-only imaging are not compared since the flat board has no interreflections.

Figures 32, 33, 35, 36 show the effect of capturing epipolar-only light for both the wedge and bowl. Since diffuse interreflections are of low frequency, high frequencies will perceive it as a constant offset [11]. Our method and each compared high frequency method can handle the constant offset, so capturing an epipolar-only image will make no difference. Low frequencies will perceive the interreflections as a phase shift instead, which causes a problem unless blocked out. Epipolar-only imaging can filter out most diffuse interreflections enabling low frequency patterns to be used in more cases.

Figures 38 and 39 show the effects of different types of diffuse interreflections. Interreflections that are localized to a certain area, but reflect strongly will have a lower number of errors, but a higher RMS. Interreflections that are prevalent everywhere, but reflect weakly will have a higher number of errors, but a lower RMS.

Under this non-ideal imaging system, for 4 to 6 images, COSU will have the best RMS, regardless if the captured images are conventional or epipolar-only. Using the unit frequency makes the set of patterns invariant to blur. The interreflections will cause a high number of incorrect pixels,

but their spread will still be near the correct match. Capturing epipolar-only images will significantly lower the number of incorrect pixels and RMS.

The 7 pattern COSN with frequency 32 or 64 shifted 3 times and the unit frequency shifted 4 times gets around the disadvantage of blurring by localizing with the unit frequency. It will provide a low RMS like COSU but also improve the number of incorrect pixels because of the high frequency pattern. These patterns are advised to be used while blocking out interreflections. The one pattern that consistently has a low RMS, even in the presence of interreflections, is the 7 pattern EMB with pattern frequencies 8, $8 + \sqrt{8}$, and 9. The unit frequency is embedded well in that pattern without losing information from the blur or being affected by the low frequency interreflections. It will still have a high number of incorrect pixels because it does not have any high frequency patterns.

8. Conclusion

We have presented new code generation and matching techniques for Structured Light 3D Scanning. By optimizing the mean-subtracted normalized cross-correlation of codes with the infinity norm metric, we can create codes that are as unique as possible for a higher number of patterns and frequencies. These codes can also handle a constant offset and can be frequency limited to work within the constraints of an imaging system.

The solution space for our problem is very large, which results in a poor optimization for a low number of patterns or few allowable frequencies. Since there is no guarantee of a global solution, starting from a random initial guess gives a consistently better solution than starting from an existing Phase Shifting code or solving in frequency space.

The mean-subtracted normalized cross-correlation matching step fits naturally with the optimization. This method was shown to work on each different set of patterns with results similar to their specialized matching method, but it does so much faster as there are just two simple steps.

Simulations showed that the occurrence of errors mostly matched with the infinity norms, although the lower infinity norm envelope patterns did not always have the lowest error and MPS did well despite its high infinity norm. This suggests that the infinity norm metric may not be the ideal metric to optimize. Real experiments corresponded to underexposed and blurred simulations, and contained hardware errors. This did not allow for our patterns to be fully tested, but did show the strength of low frequency patterns when interreflections could be filtered out.

We have shown the potential for creating codes with a low occurrence of errors, but it could be possible to improve the optimization to where a global solution could be found. Alternatively, the distance between pixels could be taken into account to create codes with a lower RMS in the match-

ing step. Results from the matching could be improved by using more than just the highest correlated code. Since the correlation ranks all codes, it could be possible to use this signature to label pixels more accurately.

References

- [1] Canon 1D Mark II Review and Sensor Analysis Plus Procedures for Evaluating Digital Camera Sensor Noise, Thermal Noise, Dynamic Range, and Full Well Capacities webpage,
<http://www.clarkvision.com/articles/evaluation-1d2/>
- [2] Episcan3D webpage,
<http://www.cs.cmu.edu/ILIM/episcan3d/html/index.html>
- [3] A. G. Bedeian, K. W. Mossholder, *On the use of coefficient of variation as a measure of diversity*, Organizational Research Methods, Volume 3, Number 3, Pages 285-297, 2000.
- [4] P. J. Green, *Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives*, Journal of the Royal Statistical Society. Series B (Methodological), Volume 46, Number 2, Pages 149-192, 1984.
- [5] M. Gupta, S. K. Nayar, *Micro Phase Shifting*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2012.
- [6] S. W. Hasinoff, F. Durand, and W. T. Freeman, *Noise-Optimal Capture for High Dynamic Range Photography*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.
- [7] E. Horn, N. Kiryati, *Toward Optimal Structured Light Patterns*, Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling, 1997.
- [8] S. Inokuchi, K. Sato, F. Matsuda, *Range imaging system for 3-D object recognition*, Proceedings of the International Conference on Pattern Recognition 1984, Pages 806-808.
- [9] M. Montilla, S. A. Orjuela-Vargas, W. Philips, *State of the art of 3D scanning systems and inspection of textile surfaces*, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE) 2014.
- [10] B. Moreno, K. Son, G. Taubin, *Embedded Phase Shifting: Robust Phase Shifting with Embedded Signals*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015.
- [11] S. K. Nayar, G. Krishnan, M. D. Grossberg, R. Raskar, *Fast Separation of Direct and Global Components of a Scene using High Frequency Illumination*, ACM Transactions on Graphics (TOG) 2006.
- [12] M. O'Toole, S. Achar, S. G. Narasimhan, K. N. Kutulakos, *Homogeneous Codes for Energy-Efficient Illumination and Imaging*, ACM SIGGRAPH 2015.
- [13] M. O'Toole, J. Mather, K. N. Kutulakos, *3D Shape and Indirect Appearance by Structured Light Transport*, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2016.
- [14] T. Pribanić, S. Mrvoš, J. Salvi, *Efficient multiple phase shift patterns for dense 3D acquisition in structured light scanning*, Image and Visual Computing, Volume 8, Issue 8, Pages 1255-1266, 2010.
- [15] J. Salvi, s. Fernandez, T. Pribanić, X. Llado, *A state of the art in structured light patterns for surface profilometry*, Pattern Recognition, Volume 43, Issue 8, Pages 2666-2680, 2010.
- [16] D. Scharstein, R. Szeliski, *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*, Journal of Computer Vision, Volume 47, Issue 1, Pages 7-42, 2002.

Naming Convention Of Patterns	
MS-NCC	Our Mean-Subtracted Normalized Cross-Correlation patterns as created in Section 5. The initial guess is a random pattern.
MS-NCC1	Our patterns as created in Section 5.4. The initial guess is the best solution within the frequency limit from MS-NCC.
COS1	A custom Phase Shifting set of patterns. The first 3 patterns are the unit frequency shifted 3 times. The remaining number of patterns are the maximum allowable frequency shifted by the corresponding number of times.
COSN	A custom Phase Shifting set of patterns. Similar to COS1, two frequencies are used. The first is shifted 3 times and the second is shifted by the remaining number of patterns times. The two frequencies are chosen such that one frequency is the maximum allowable, and the other is the frequency that minimizes the infinity norm of the correlation matrix, as described in Section 5. The both orderings of the frequencies are taken into account.
COSU	A Phase Shifting set of patterns where the unit frequency is shifted by the number of patterns.
MPS	Micro Phase Shifting [5]. The first frequency is shifted 3 times for the first 3 patterns. For each of the remaining patterns, a new unshifted frequency is projected. The range of frequencies that are searched through are $\frac{8}{9}$ to 1 times the maximum allowable frequency.
EMB	Embedded Phase Shifting [10]. For 5 patterns, the embedded frequencies are the maximum allowable and 1. The first pattern frequency is shifted 3 times and the second pattern frequency is shifted 2 times. For 6 patterns, everything is the same except that the second pattern frequency is shifted 3 times. For 7 patterns, the embedded frequencies are the maximum allowable, the square root of the maximum allowable, and 1. The first pattern frequency is shifted 3 times and the remaining two frequencies are shifted twice each.
Naming Convention Of Matching Algorithms	
MS-NCC	Our Mean-Subtracted Normalized Cross-Correlation code matching as described in Section 4. This matching algorithm can be applied to any pattern.
SPEC	The matching algorithm used is the Structured Light pattern's specialized algorithm.

Table 1. The abbreviations and pattern descriptions of each Structured Light pattern used.

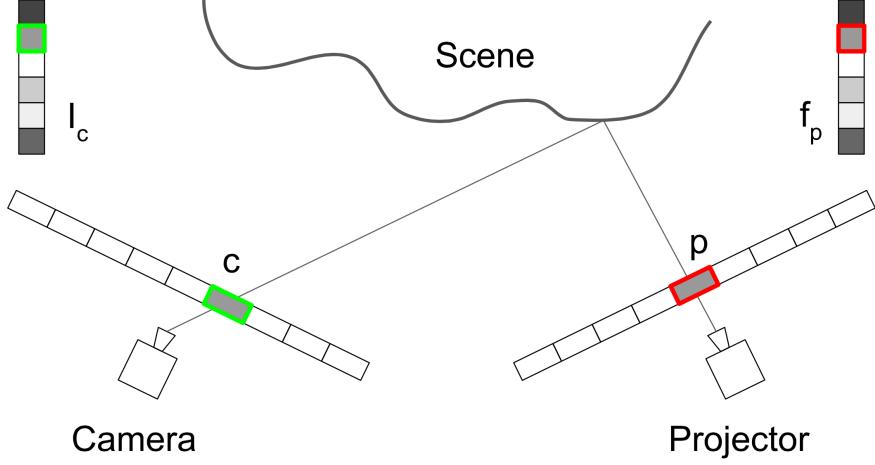


Figure 1. The transmission of light from projector pixels to camera pixels. Light that leaves the center of projection of the projector through pixel p will be reflected off the scene and the camera will capture the light that goes towards its center of projection. Pixel c that the light passes through will be the pixel corresponding to p . Transmitting a code vector, f_p , involves projecting one pattern for each code term and capturing an image. In this example, the second term is transmitted and a total of 6 images would be needed to capture I_c . Although only one active pixel is shown, in practice all pixels will be projecting and capturing at the same time.

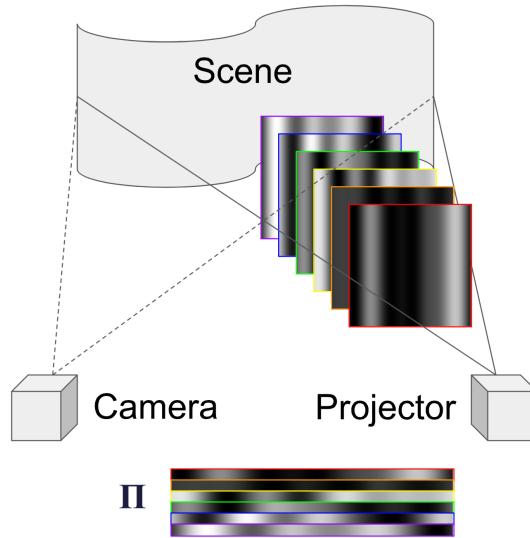


Figure 2. Transmitting an $M \times N$ Π matrix as a series of 2D projection patterns. The columns of Π correspond to M -length Structured Light codes, as shown in Figure 1. The rows of Π correspond to N -length projection patterns. Each row of Π is repeated to form a 2D projection pattern. M 2D projection patterns must be projected and captured to transmit the Π matrix.

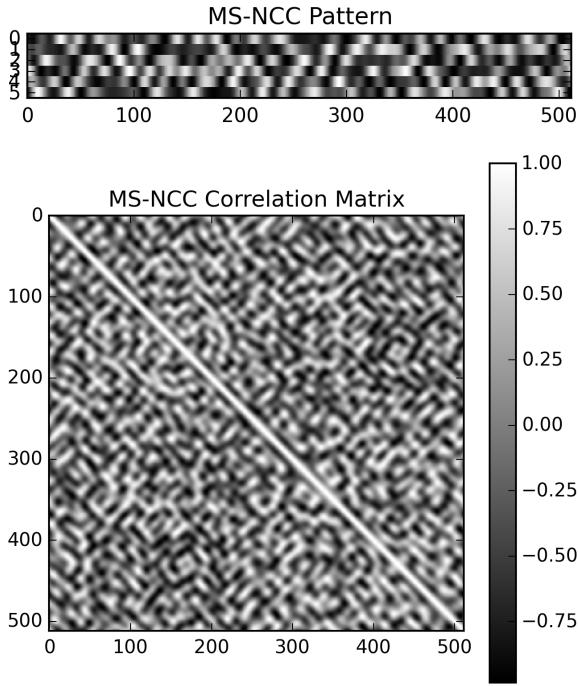


Figure 3. Top: The projection matrix, Π , resulting from optimizing with MS-NCC under the constraints of using 6 patterns, 512 codes, and a maximum allowable frequency of 32. Bottom: The corresponding normalized correlation matrix, $\mathbf{Q}_{\text{norm}}^T \mathbf{Q}_{\text{norm}}$.

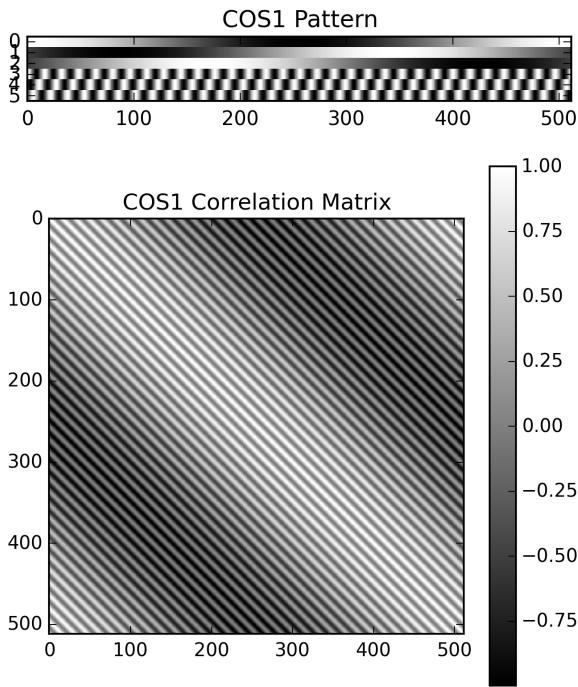


Figure 4. Top: The projection matrix created from using a unit cosine and a frequency 32 cosine both shifted three times (COS1). Bottom: The corresponding correlation matrix.

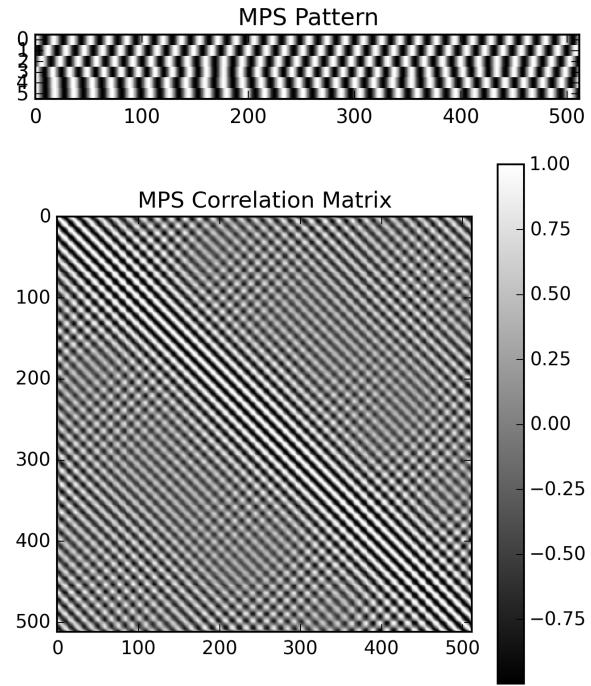


Figure 5. Top: The projection matrix created from using MPS with a 6 patterns, 32 frequency limit. Bottom: The corresponding correlation matrix.

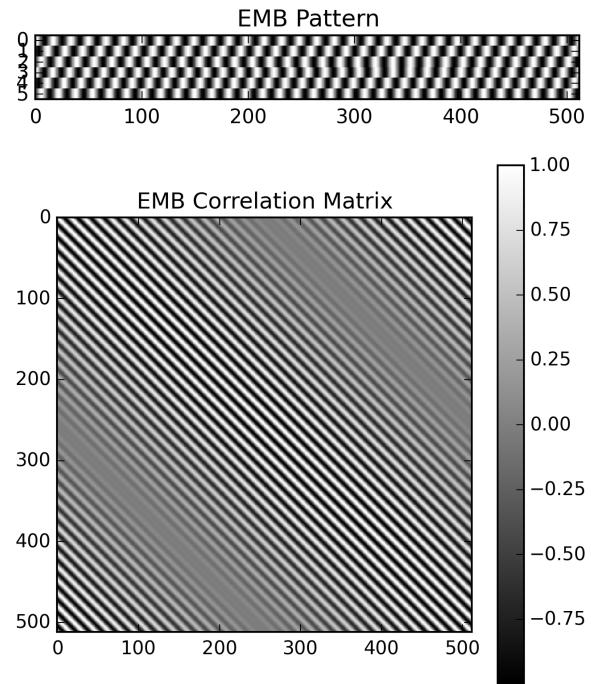


Figure 6. Top: The projection matrix created from using EMB with embedded frequencies 31 and 1, or equivalently, pattern frequencies 31 and 32. Both cosines are shifted three times. Bottom: The corresponding correlation matrix.

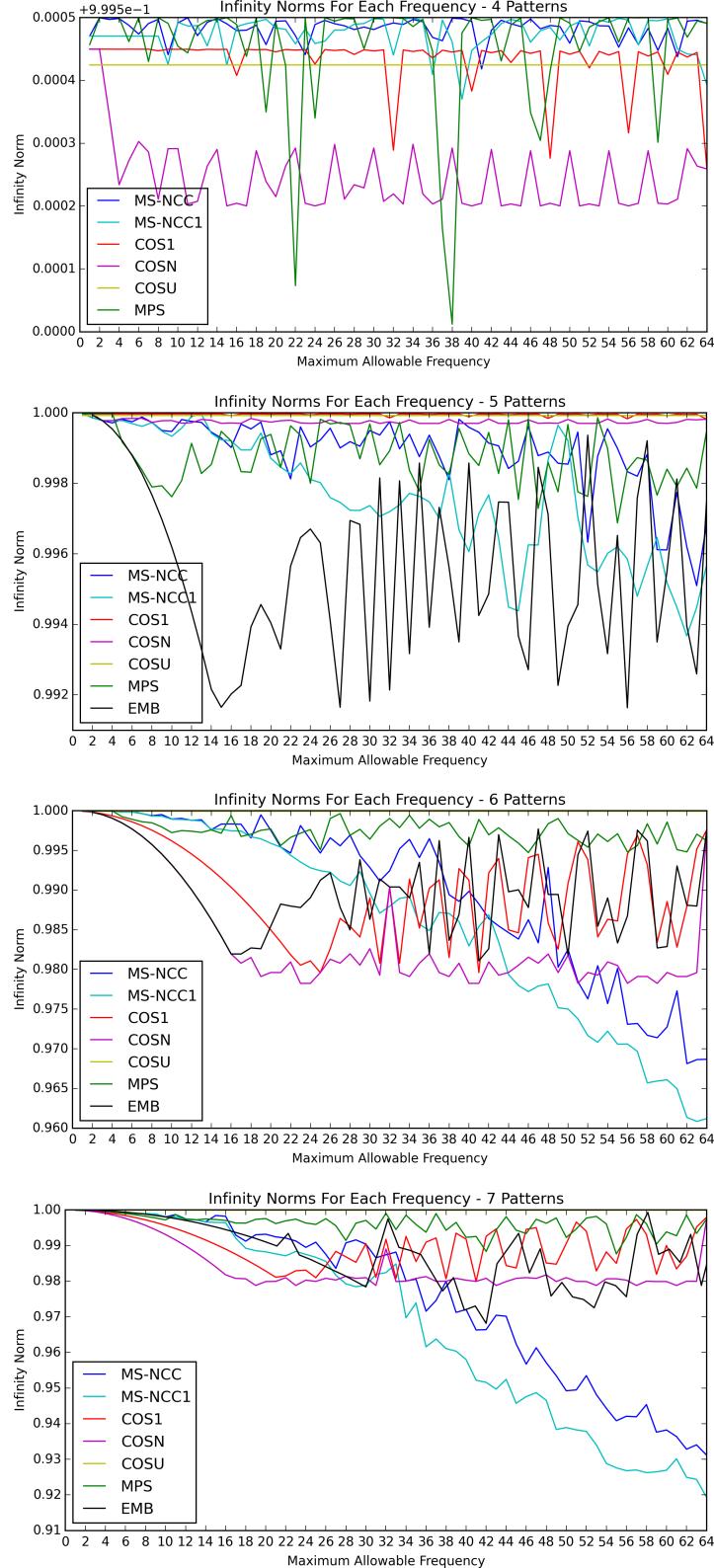


Figure 7. $\|\mathbf{Q}_{\text{norm}}^T \mathbf{Q}_{\text{norm}} - \mathbf{I}\|_\infty$ for each type of pattern and each maximum allowable frequency from 1 to 64 (each pattern is 512 pixels long). The norms displayed are in fact the infinity norms, not the 100-norm. Top to Bottom: The infinity norm values are shown for 4, 5, 6, and 7 patterns (the number of rows in Π). The labels follow the naming conventions of Table 1. See Figures 8 and 9 for the lower envelopes of the infinity norms for each pattern.

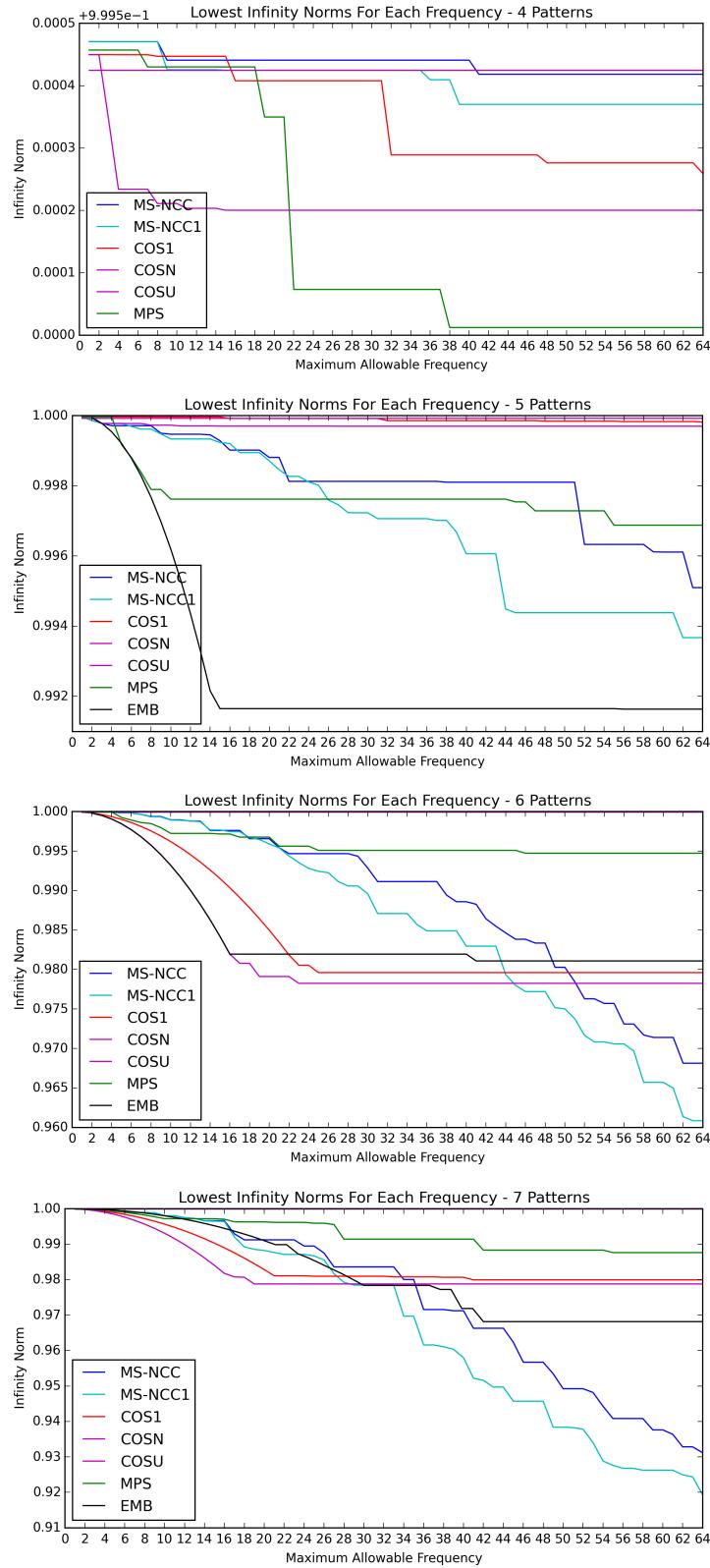


Figure 8. The lower envelope of the infinity norm for each pattern. Top to Bottom: The infinity norm values are shown for 4, 5, 6, and 7 patterns.

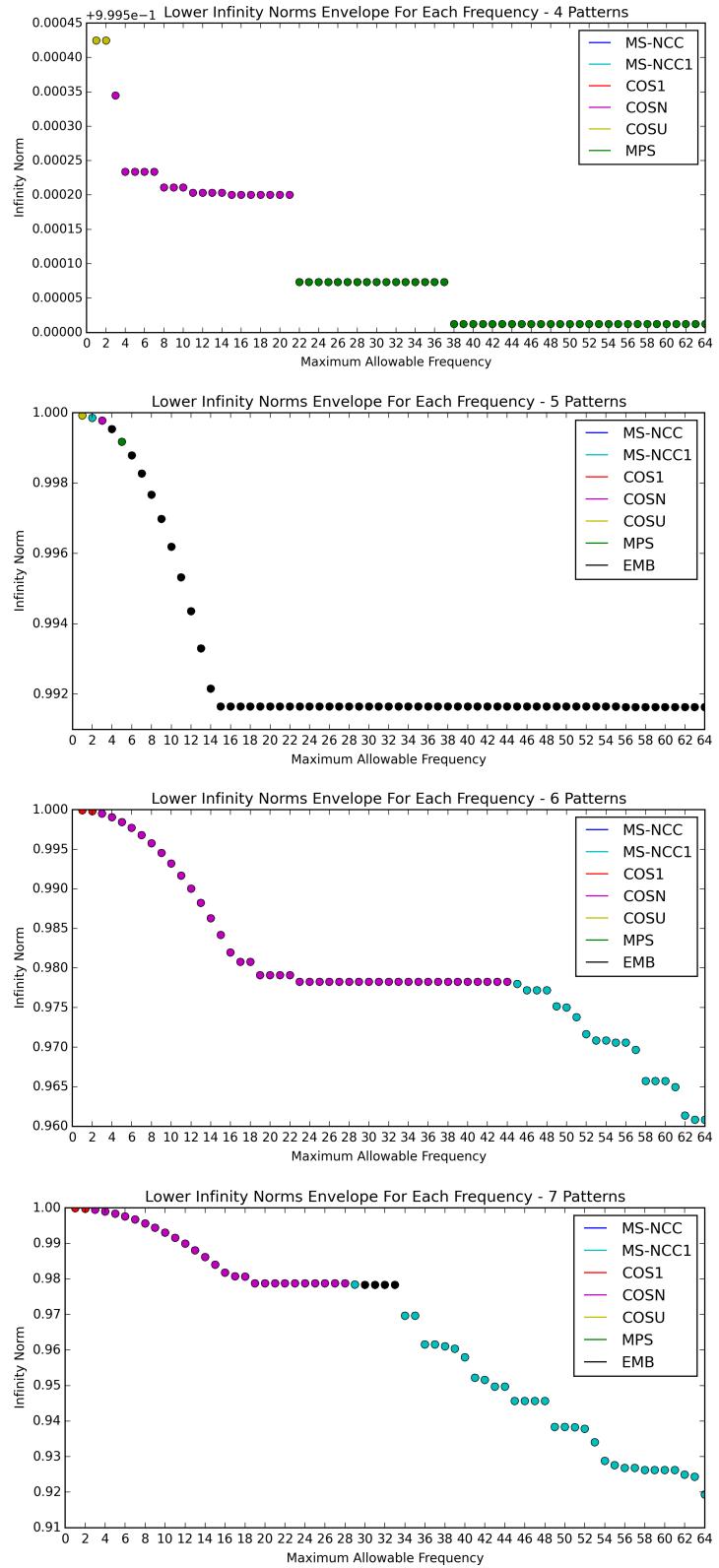


Figure 9. The pattern with the lowest infinity norm up to a given frequency is shown at that frequency. Top to Bottom: The infinity norm values are shown for 4, 5, 6, and 7 patterns.

Frequency	Domain	Length of Patterns in Pixels		
		8	16	32
1	Spatial	Time: 0.24 Norm: 0.8918	-	-
1	Frequency	Time: 0.02 Norm: 0.9717	-	-
2	Spatial	Time: 0.55 Norm: 0.6505	Time: 2.02 Norm: 0.9015	-
2	Frequency	Time: 0.10 Norm: 0.6502	Time: 0.01 Norm: 0.9963	-
4	Spatial	- -	Time: 1.64 Norm: 0.7564	Time: 6.49 Norm: 0.9148
4	Frequency	- -	Time: 1.18 Norm: 0.7535	Time: 0.02 Norm: 0.9991
6	Spatial	- -	-	Time: 5.91 Norm: 0.8528
6	Frequency	- -	-	Time: 0.02 Norm: 0.9923
8	Spatial	- -	-	Time: 6.86 Norm: 0.8388
8	Frequency	- -	-	Time: 2.79 Norm: 0.8444

Table 2. The infinity norms and the time it took in minutes to optimize the MS-NCC codes in the spatial domain and frequency domain for various pattern lengths, maximum allowable frequencies, and a code length of 6. In the spatial domain, each optimization ran for many iterations and came up with a solution. In the frequency domain, only once the maximum allowable frequency was at least one fourth of the number of pixels did the optimization run. Otherwise, the optimization would only iterate a single time. In the cases that the optimization ran in both domains, the resulting infinity norm was very similar, but took less time in the frequency domain.

Num Pixels, Max Frequency	Initial Guess		
	Random	Phase Shifting	MPS
16 , 4	0.6081	0.5623	0.6112
32 , 4	0.8131	0.7645	0.7697
64 , 4	0.9379	0.9300	0.9313
128 , 4	0.9854	0.9850	0.9890
256 , 4	0.9985	0.9962	-
256 , 8	0.9964	0.9866	-
512 , 8	0.9994	0.9967	-
512 , 24	0.9964	0.9869	-
512 , 42	0.9864	0.9958	-
512 , 46	0.9863	0.9874	-

Table 3. The infinity norms of the results of the optimization when using initial guesses of random values, general Phase Shifting codes, and MPS codes. The norms are shown for a code length of 6 and various pattern lengths and maximum allowable frequencies. Starting with MPS shows improvements for small pattern lengths but was not guaranteed to be better. Starting from Phase Shifting codes give a lower infinity norm at all pattern sizes. These codes are in or near local minima, which means the optimization runs for only a few steps but the resulting patterns are very similar to a sinusoid and may not be the best solution, as shown for higher maximum allowable frequencies.

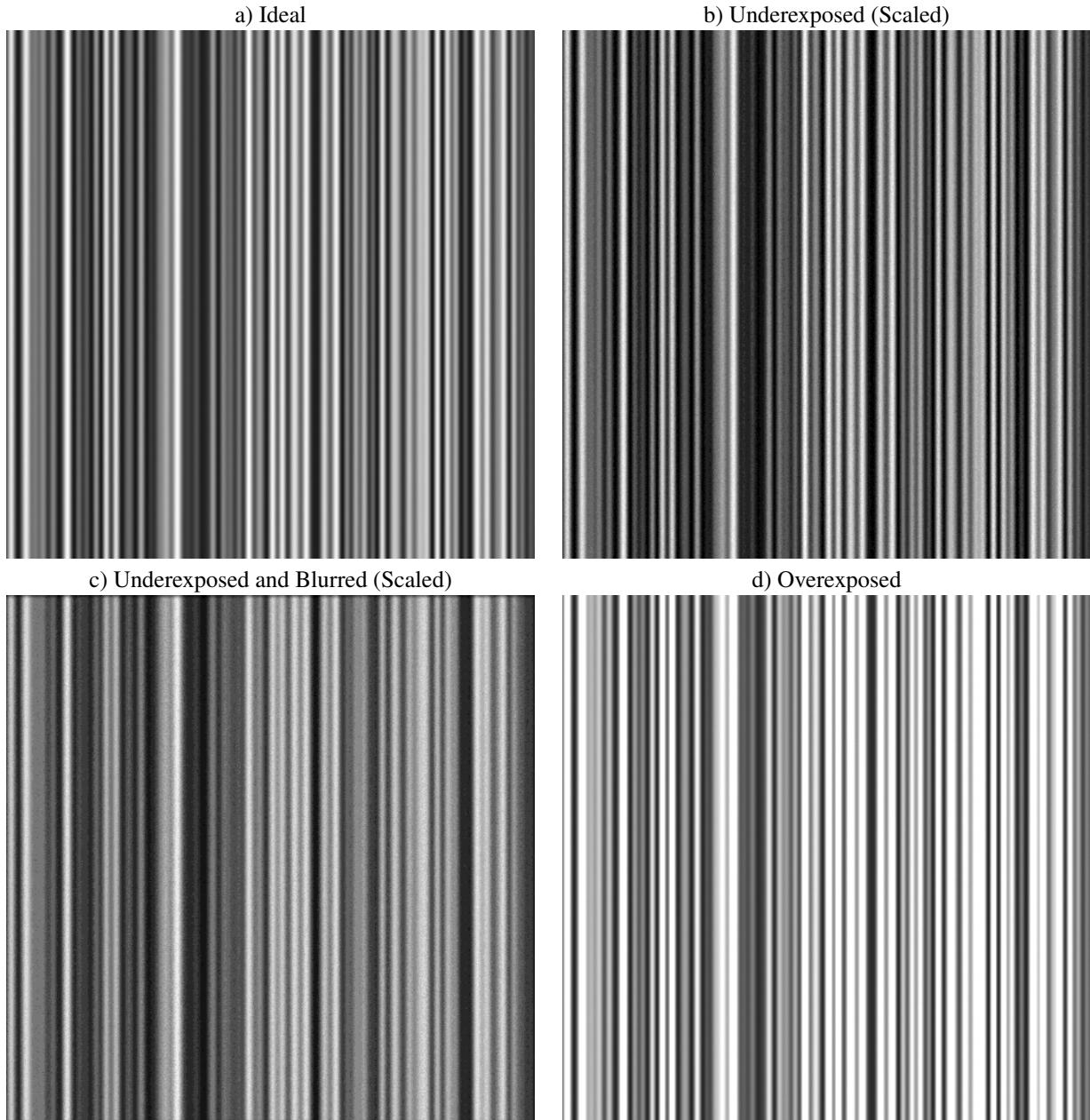


Figure 10. Examples of simulated images. The images are of one pattern of a 7-pattern MS-NCC set with a maximum frequency of 64. a) An ideal image with the best exposure and no blur. b) An underexposed image scaled up by 16 times to visually match the ideal exposure. Notice the higher presence of noise. c) An underexposed and blurred (8-pixel circular convolution) image scaled up by 16 times to visually match the ideal exposure. Notice the attenuation of the high frequency content. d) An overexposed (by 1.5 times) image where many pixels have hit the full well capacity. These pixels will have an incorrect amplitude that will lead to errors in the matching.

Simulated Images, Ideal Exposure, Percentage of Incorrect Pixels

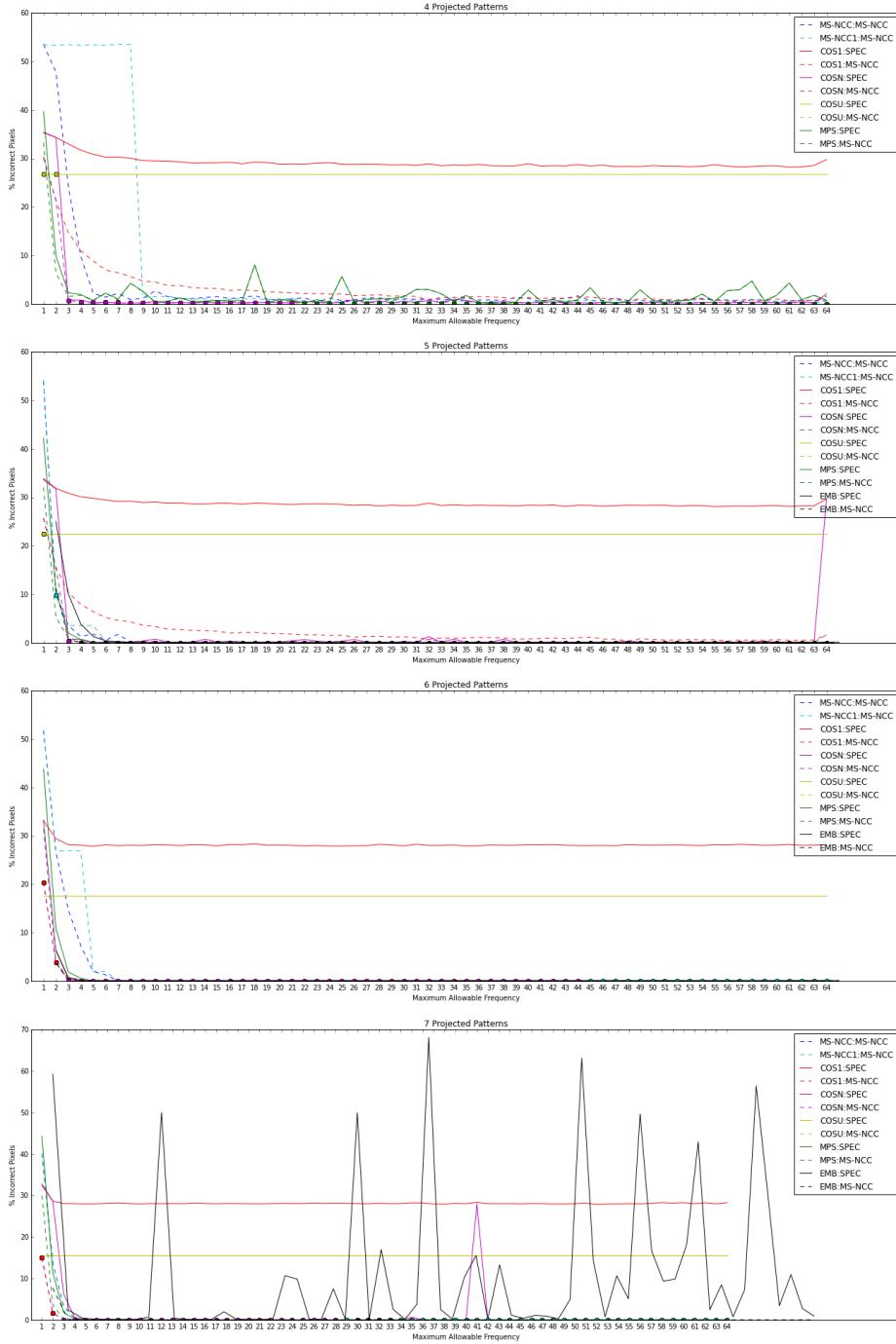


Figure 11. The percentage of incorrect pixels for each pattern and maximum allowable frequency of a simulated image of an iso-depth plane under ideal exposure. Each of the patterns shown in Figure 7 are shown here. A solid line indicates the matching was done with the pattern's specialized method, while a dashed line indicates that MS-NCC was applied. The discrete dots represent the errors for the patterns forming the lower infinity norms envelope of Figure 9, matched with MS-NCC. It is important to note that the ground truth for each pixel is a discrete value. This makes noise more detrimental to the continuous matching. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The full range of errors are shown. See Figure 12 for a zoomed in version to distinguish between the methods that have the lowest error. The low noise level causes very few errors, except for very low frequencies which have poor precision. The lower infinity norm envelope patterns generally have the lowest errors. Note: The high errors for EMB for 7 patterns arises from the continuous matching method. Two of the three relative phases are found with only two shifted patterns, not three. This is highly prone to error and the absolute phase gotten from the two relative phases will likely not be accurate.

Simulated Images, Ideal Exposure, Percentage of Incorrect Pixels, Zoomed

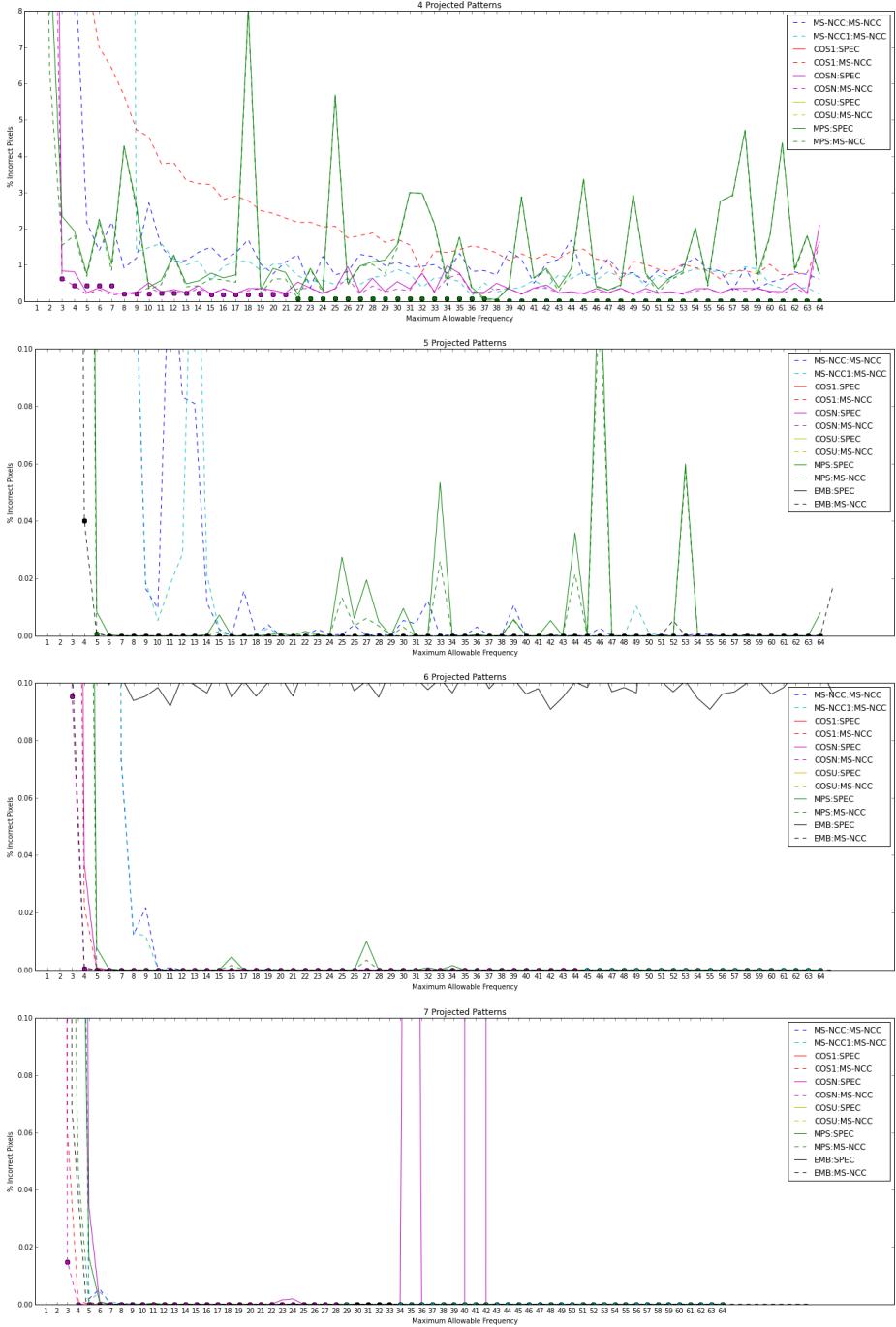


Figure 12. The zoomed in version of Figure 11 to distinguish between the methods that have the lowest error.

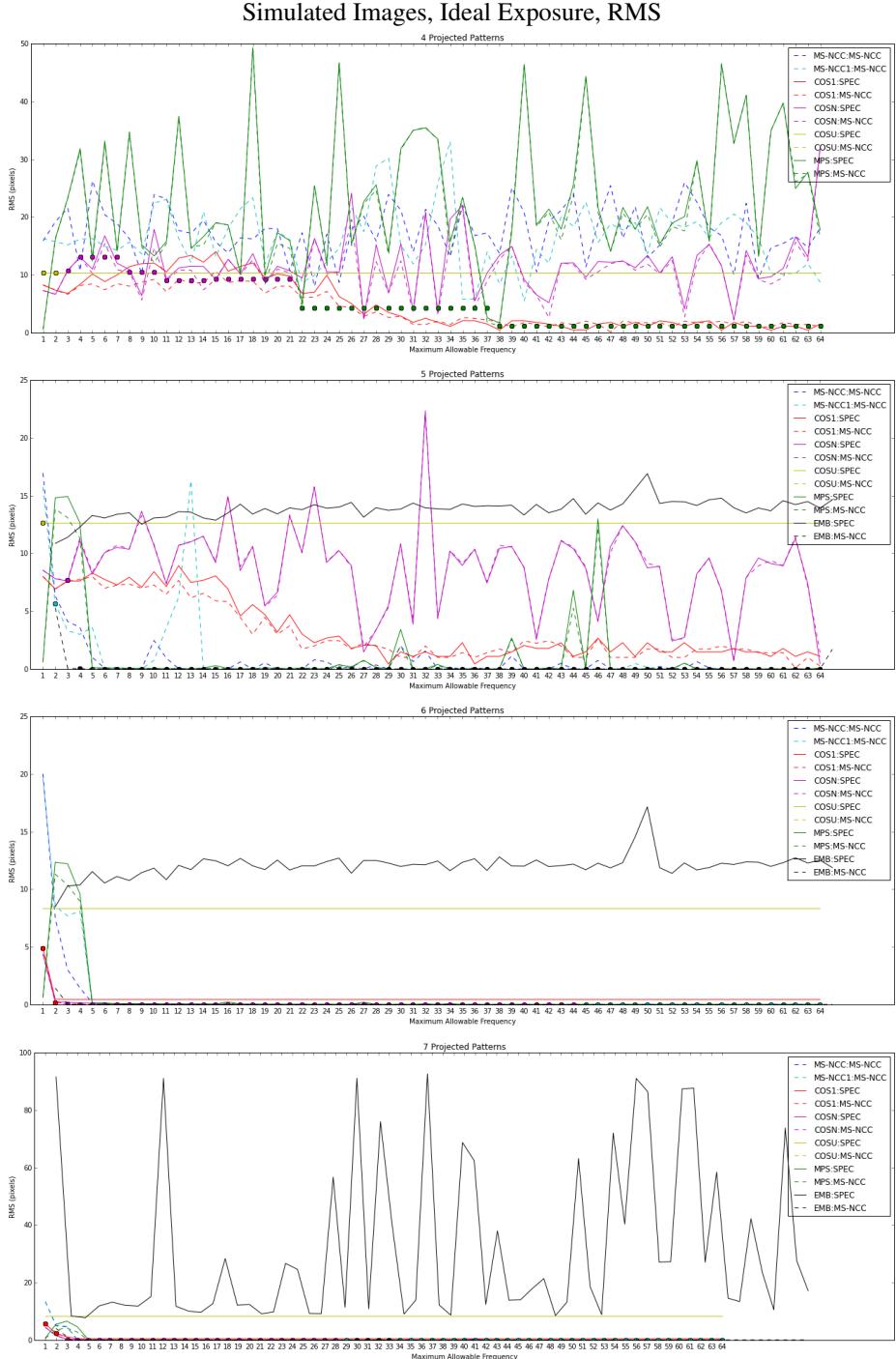


Figure 13. The RMS (in pixels) for each pattern and maximum allowable frequency of a simulated image of an iso-depth plane under ideal exposure. Solid: Specialized matching method. Dashed: MS-NCC matching. The discrete dots represent the errors for the patterns forming the lower infinity norms envelope of Figure 9, matched with MS-NCC. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The full range of errors are shown. See Figure 14 for a zoomed in version to distinguish between the methods that have the lowest error.

Simulated Images, Ideal Exposure, RMS, Zoomed

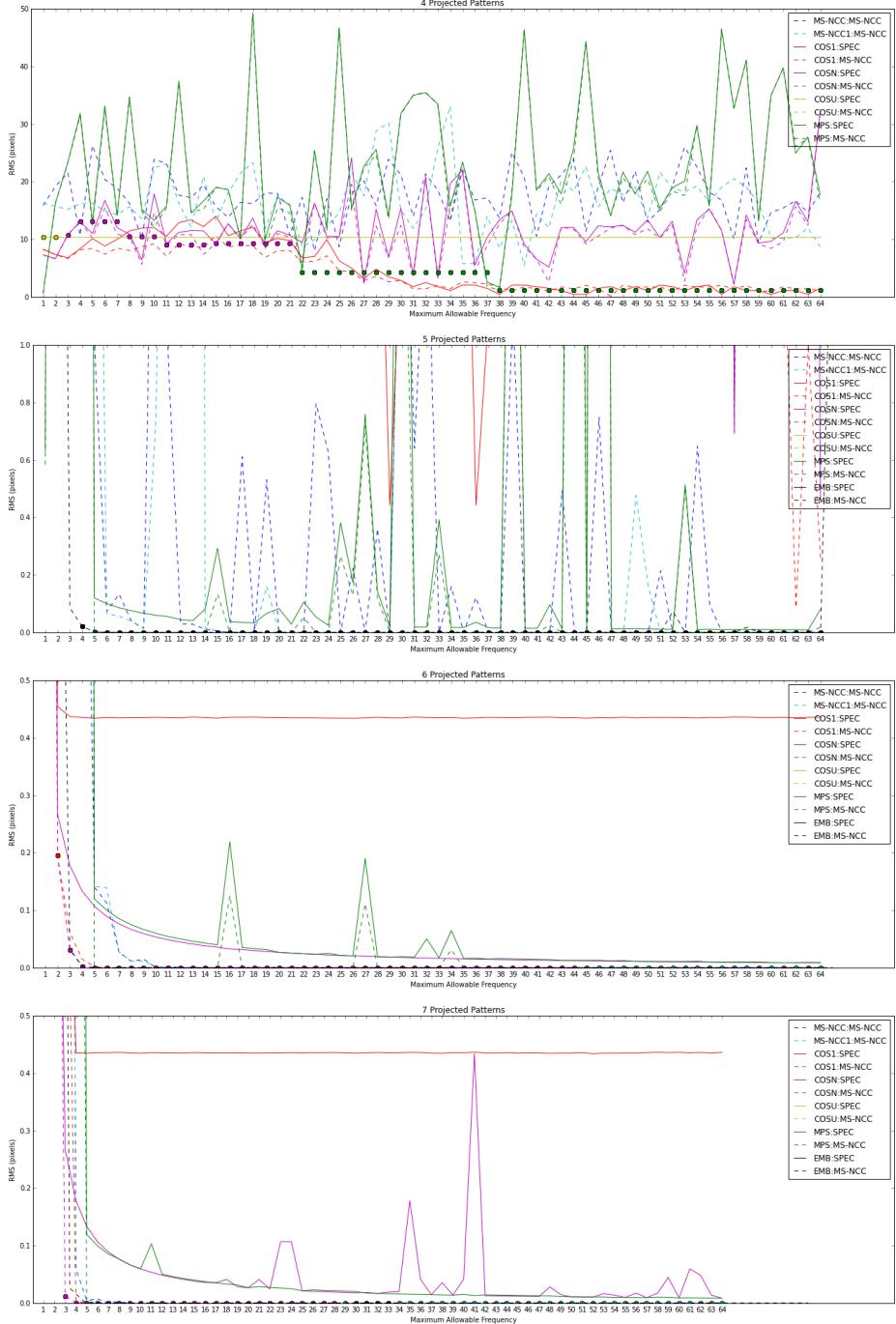


Figure 14. The zoomed in version of Figure 13 to distinguish between the methods that have the lowest error.

Simulated Images, Underexposure, Percentage of Incorrect Pixels

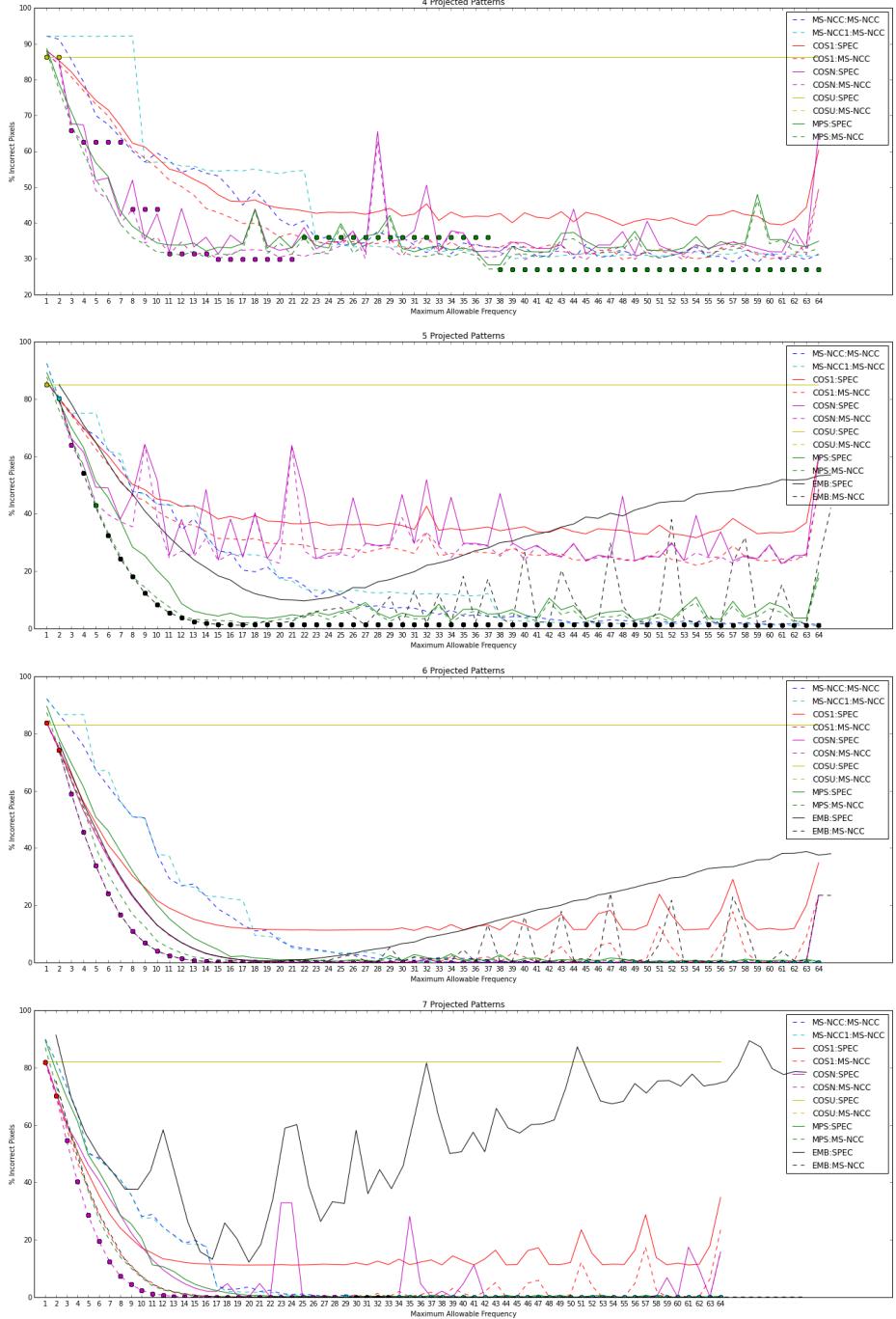


Figure 15. The percentage of incorrect pixels for each pattern and maximum allowable frequency of a simulated image of an iso-depth plane with four stops of underexposure. Solid: Specialized matching method. Dashed: MS-NCC matching. The discrete dots represent the errors for the patterns forming the lower infinity norms envelope of Figure 9, matched with MS-NCC. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The full range of errors are shown. See Figure 16 for a zoomed in version to distinguish between the methods that have the lowest error. The errors mostly match with the infinity norms of each pattern, except for MPS, which has poor infinity norms, but good error performance. As in the ideal case, the lower infinity norm envelope patterns generally have the lowest errors, apart from the 4 pattern case where the small differences between the norms are overshadowed by the noise.

Simulated Images, Underexposure, Percentage of Incorrect Pixels, Zoomed

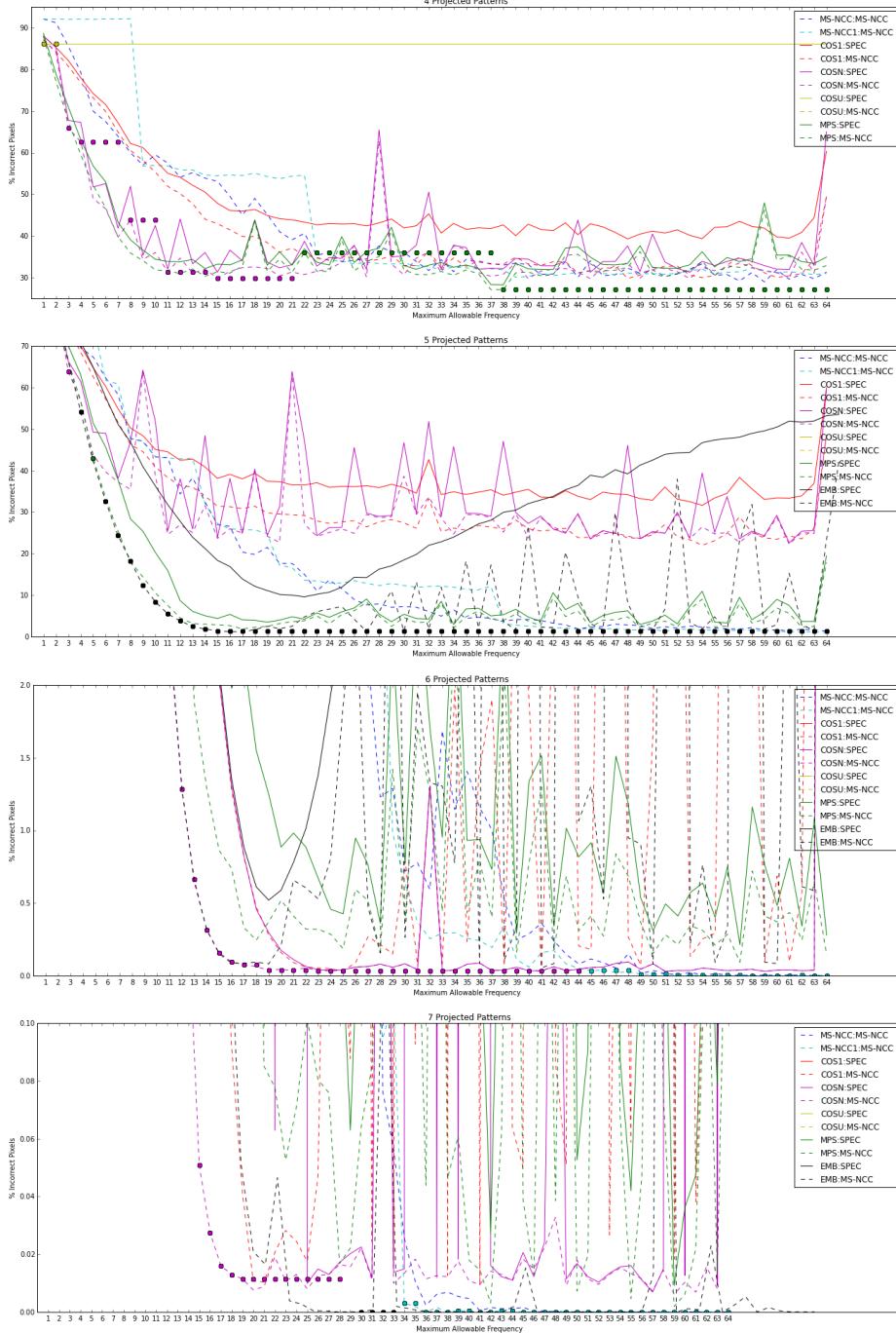


Figure 16. The zoomed in version of Figure 15 to distinguish between the methods that have the lowest error.

Simulated Images, Underexposure, RMS

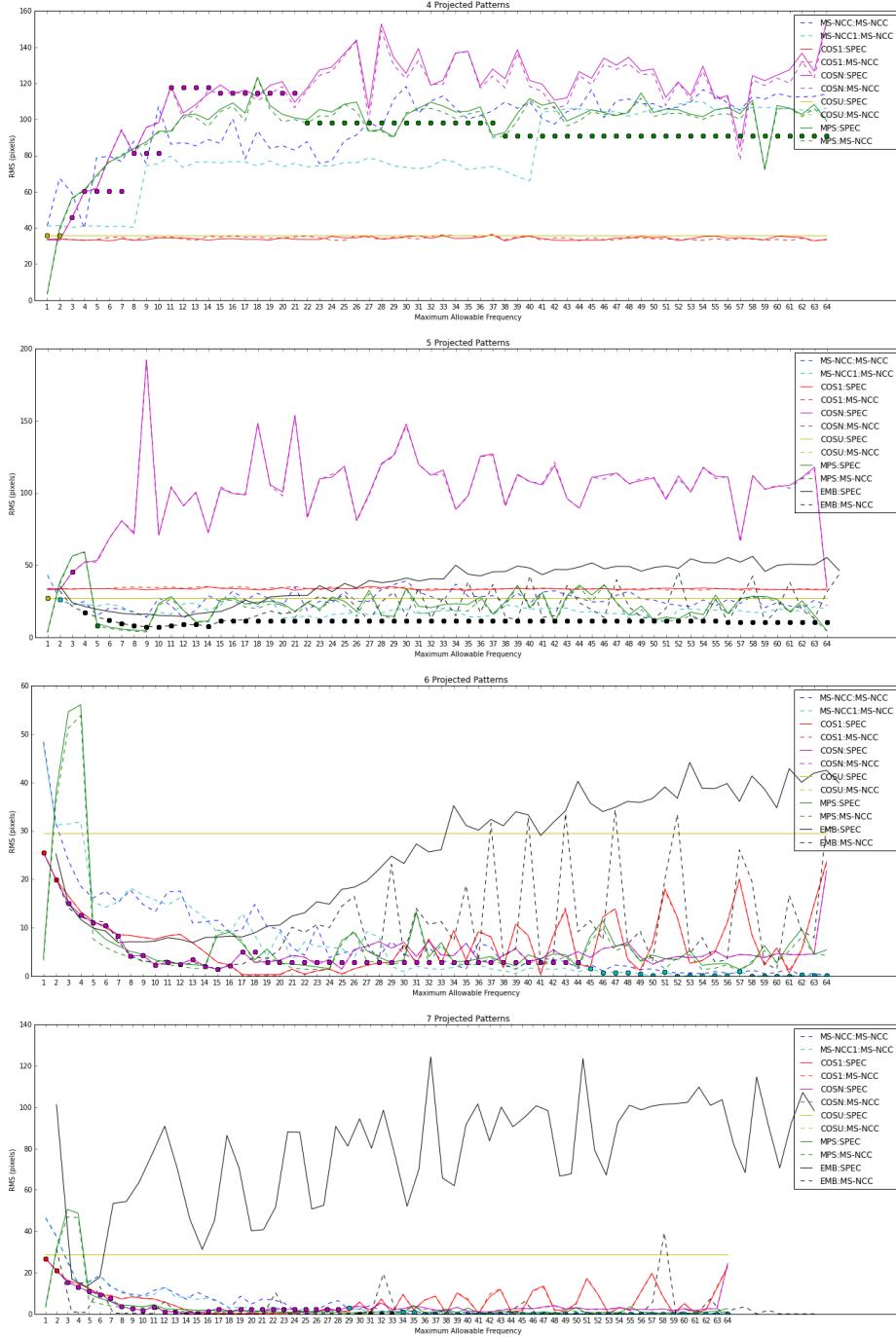


Figure 17. The RMS (in pixels) for each pattern and maximum allowable frequency of a simulated image of an iso-depth plane with four stops of underexposure. Solid: Specialized matching method. Dashed: MS-NCC matching. The discrete dots represent the errors for the patterns forming the lower infinity norms envelope of Figure 9, matched with MS-NCC. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The full range of errors are shown. See Figure 18 for a zoomed in version to distinguish between the methods that have the lowest error. Patterns with few incorrect pixels also have low RMS apart for very low frequency patterns that benefit from a lower level of confusion between codes whose pixels are far from each other. The lower infinity norm envelope patterns do not always have the best RMS since they are not optimized for that.

Simulated Images, Underexposure, RMS, Zoomed

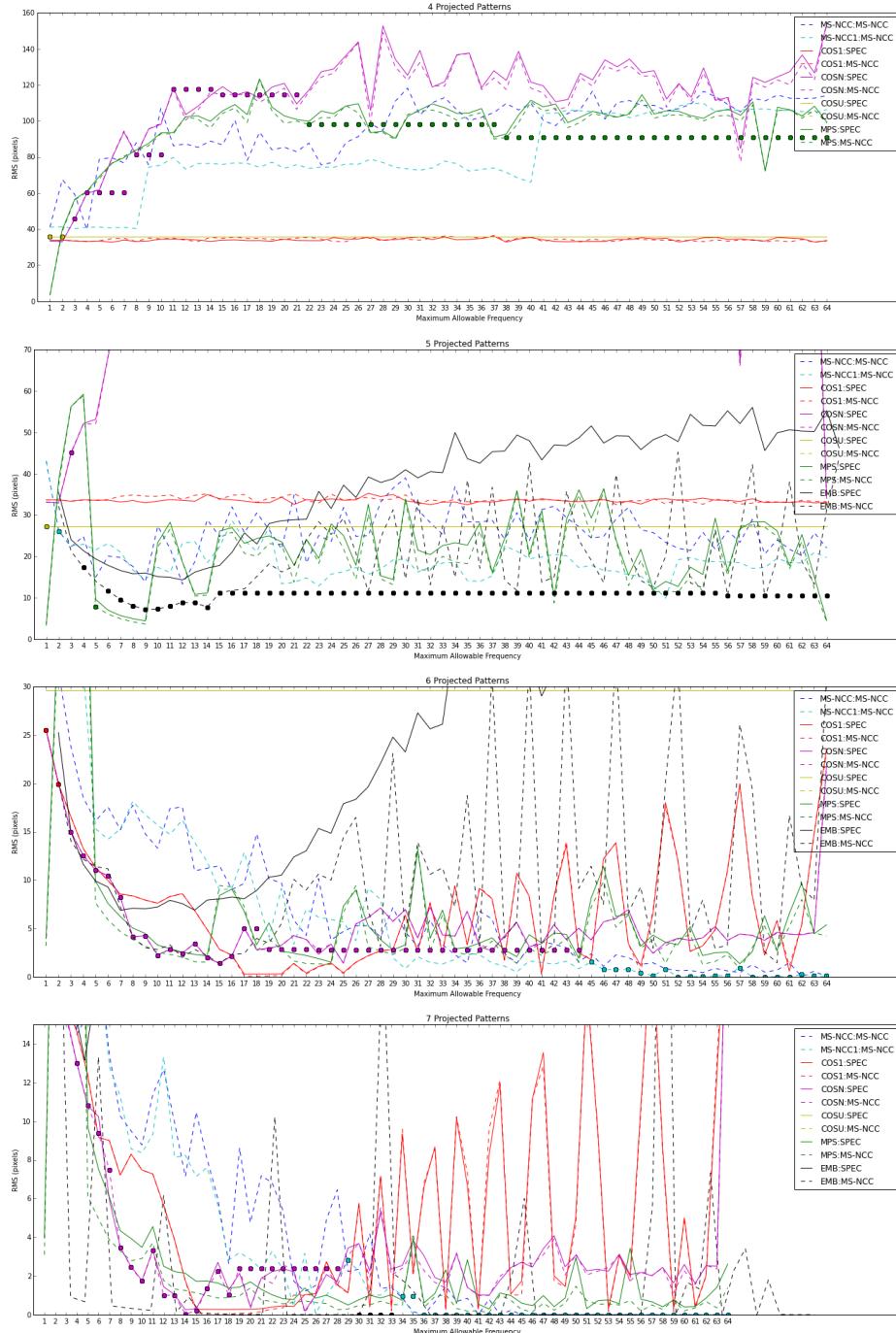


Figure 18. The zoomed in version of Figure 17 to distinguish between the methods that have the lowest error.

Simulated Images, Underexposure and Blur, Percentage of Incorrect Pixels

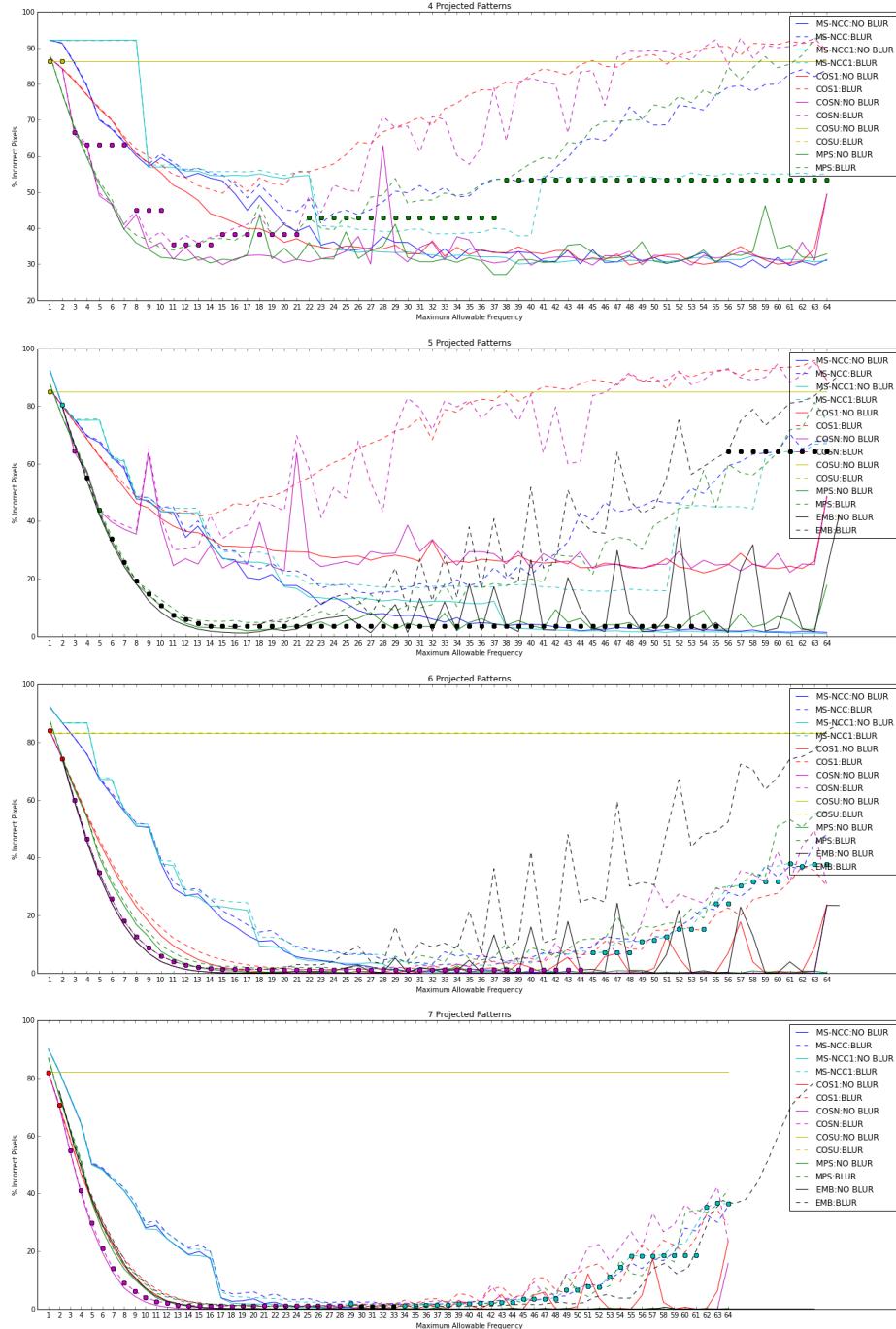


Figure 19. The percentage of incorrect pixels in the presence of blur for each pattern and maximum allowable frequency of a simulated image of an iso-depth plane with four stops of underexposure. Solid: No blur. MS-NCC matching. Dashed: 8 pixel circular blur. MS-NCC matching. The discrete dots represent the errors for the patterns forming the lower infinity norms envelope of Figure 9 in the presence of blur. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The errors of the low frequencies remain the same, as high frequencies have much higher errors.

Simulated Images, Underexposure and Blur, RMS

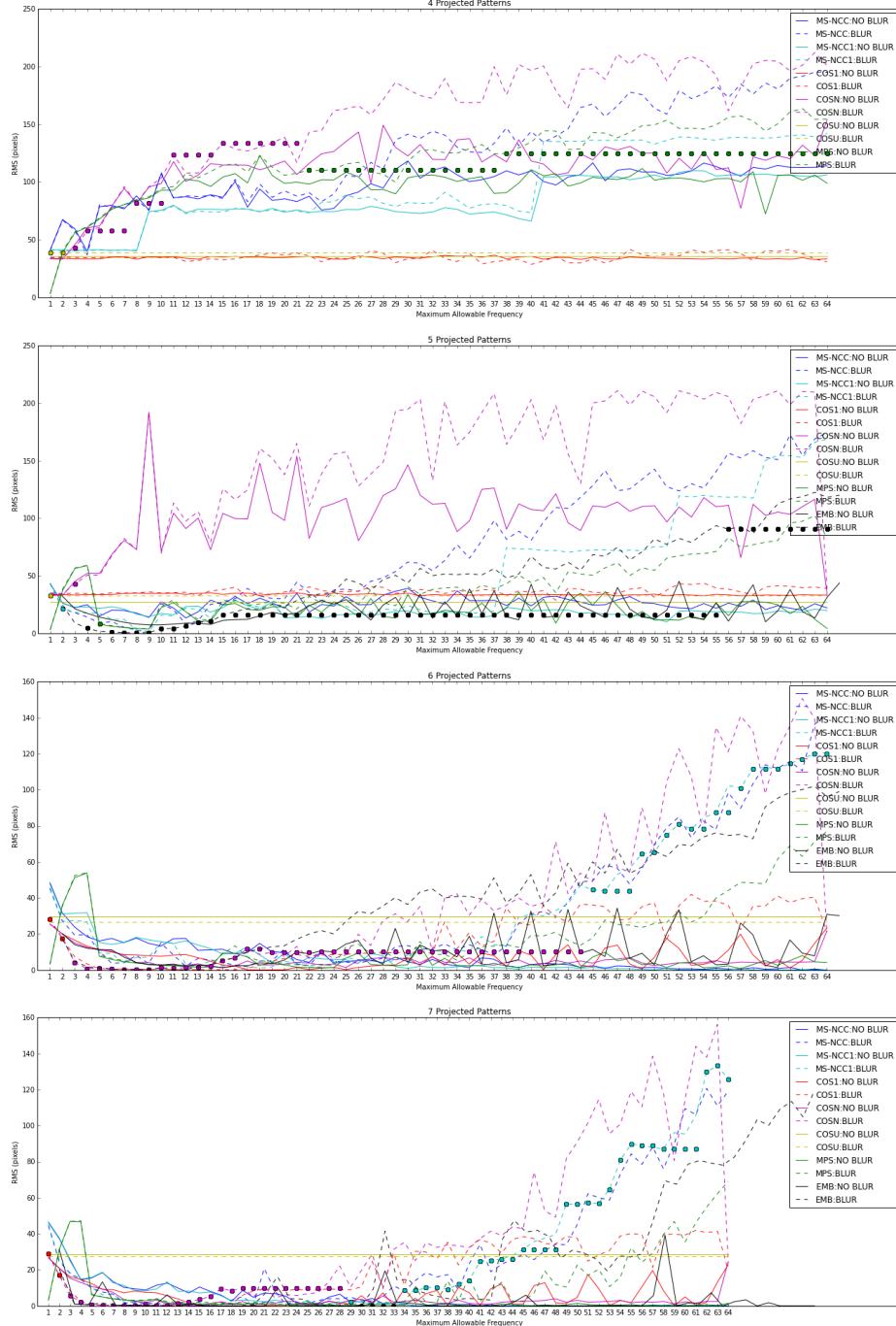


Figure 20. The RMS (in pixels) in the presence of blur for each pattern and maximum allowable frequency of a simulated image of an iso-depth plane with four stops of underexposure. Solid: No blur. NS-NCC matching. Dashed: 8 pixel circular blur. MS-NCC matching. The discrete dots represent the errors for the patterns forming the lower infinity norms envelope of Figure 9 in the presence of blur. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The errors of the low frequencies remain the same, as high frequencies have much higher errors.

Simulated Images, Overexposure, Percentage of Incorrect Pixels

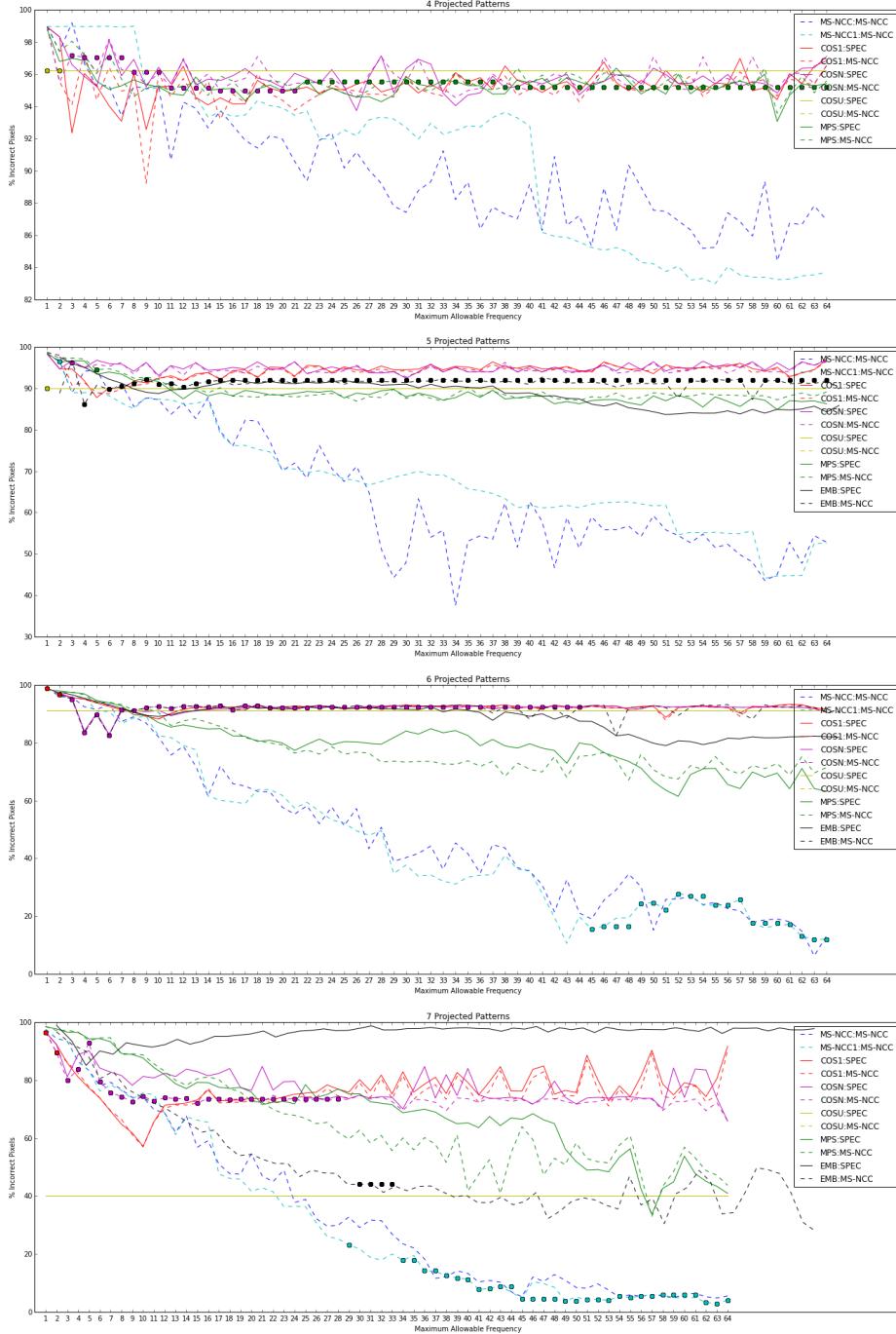


Figure 21. The percentage of incorrect pixels for each pattern and maximum allowable frequency of a simulated image of an iso-depth plane at 1.5x overexposure. Solid: Specialized matching method. Dashed: MS-NCC matching. The discrete dots represent the errors for the patterns forming the lower infinity norms envelope of Figure 9, matched with MS-NCC. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The MS-NCC codes have a fewer number of errors than all other methods when overexposed.

Simulated Images, Overexposure, RMS

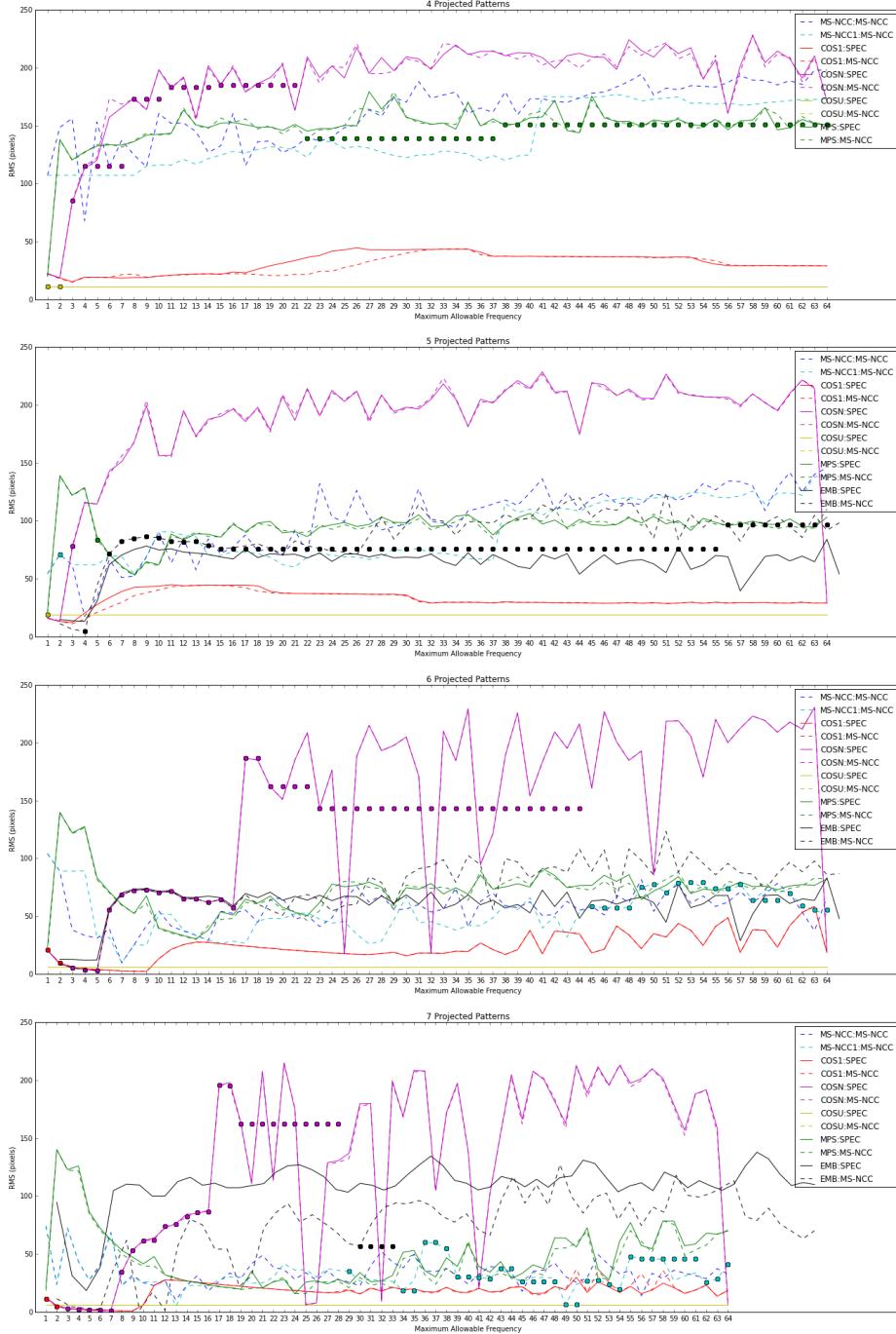


Figure 22. The RMS (in pixels) for each pattern and maximum allowable frequency of a simulated image of an iso-depth plane at 1.5x overexposure. Solid: Specialized matching method. Dashed: MS-NCC matching. The discrete dots represent the errors for the patterns forming the lower infinity norms envelope of Figure 9, matched with MS-NCC. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The trend in the RMS is mostly unchanged except for the unit cosine pattern that does best.

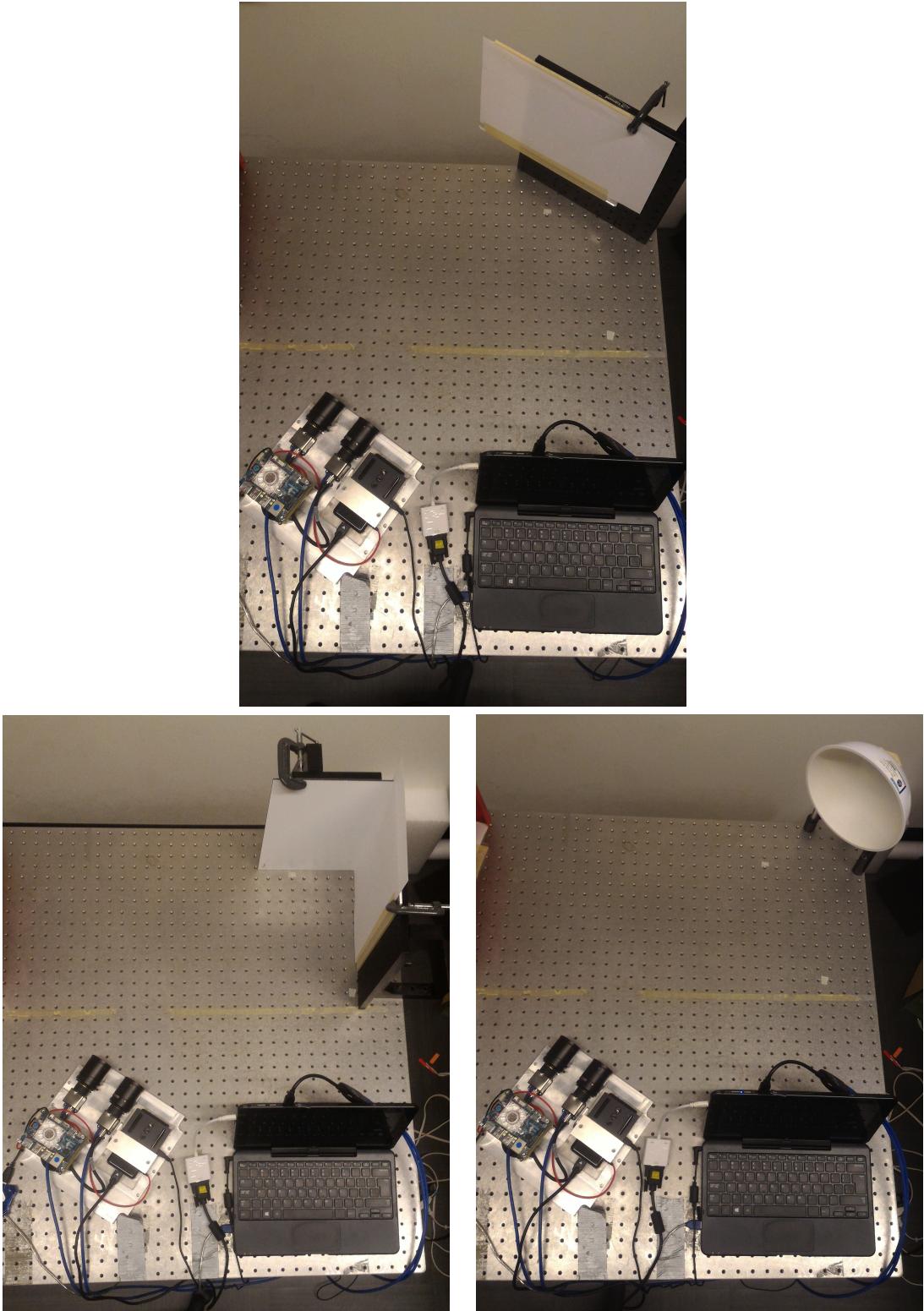


Figure 23. We use the Episcan3D [2] system to take both conventional and epipolar-only images of our targets. Although the system has two cameras, which can be used for stereo, only one is used. The single flat board and the two boards that make up the wedge are 8 by 12 inches in size and 24 inches from the system. The bowl has a 9 inch diameter, is 3.5 inches deep, and 27 inches from the system. Each surface is diffuse, so the wedge and bowl will produce diffuse interreflections. Please zoom in for details.

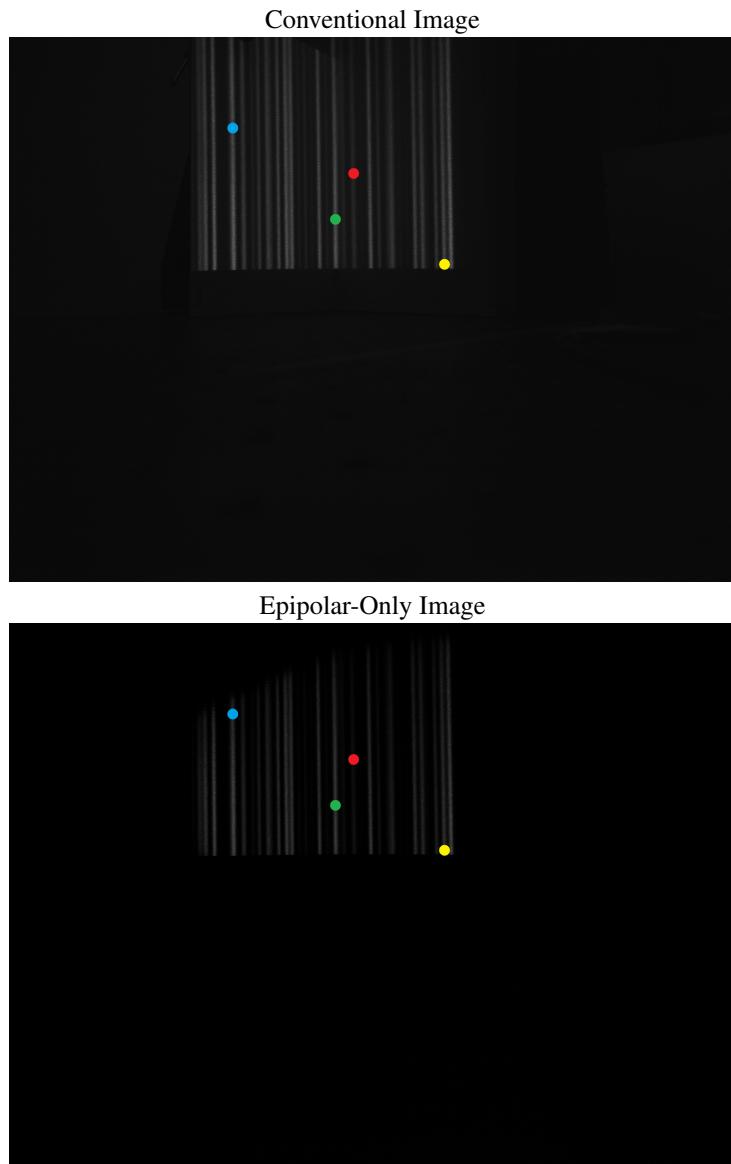


Figure 24. One pattern of a 7-pattern MS-NCC set with a maximum frequency of 40 was continuously projected and 500 conventional and epipolar-only images were captured. An example image from each imaging technique is shown. 4 pixels of differing intensity and location within the epipolar-only region are marked. Their intensity evolution through time is shown in Table 4 and Figure 25.

Pixel	Imaging	Mean Intensity	Standard Deviation	Coefficient of Variation
1 (Blue)	Conventional	77.66	2.945	0.0379
1 (Blue)	Epipolar-Only	45.00	2.776	0.0631
2 (Green)	Conventional	80.31	3.370	0.0420
2 (Green)	Epipolar-Only	56.07	2.905	0.0518
3 (Red)	Conventional	34.47	1.750	0.0506
3 (Red)	Epipolar-Only	10.65	1.082	0.1016
4 (Yellow)	Conventional	68.19	2.259	0.0331
4 (Yellow)	Epipolar-Only	48.21	1.867	0.0387

Table 4. The mean intensity, standard deviation, and coefficient of variation [3] are shown for each labeled pixel in Figure 24 through 500 images. The coefficient of variation is less for conventional imaging, so there will be lower errors due to intensity variation.

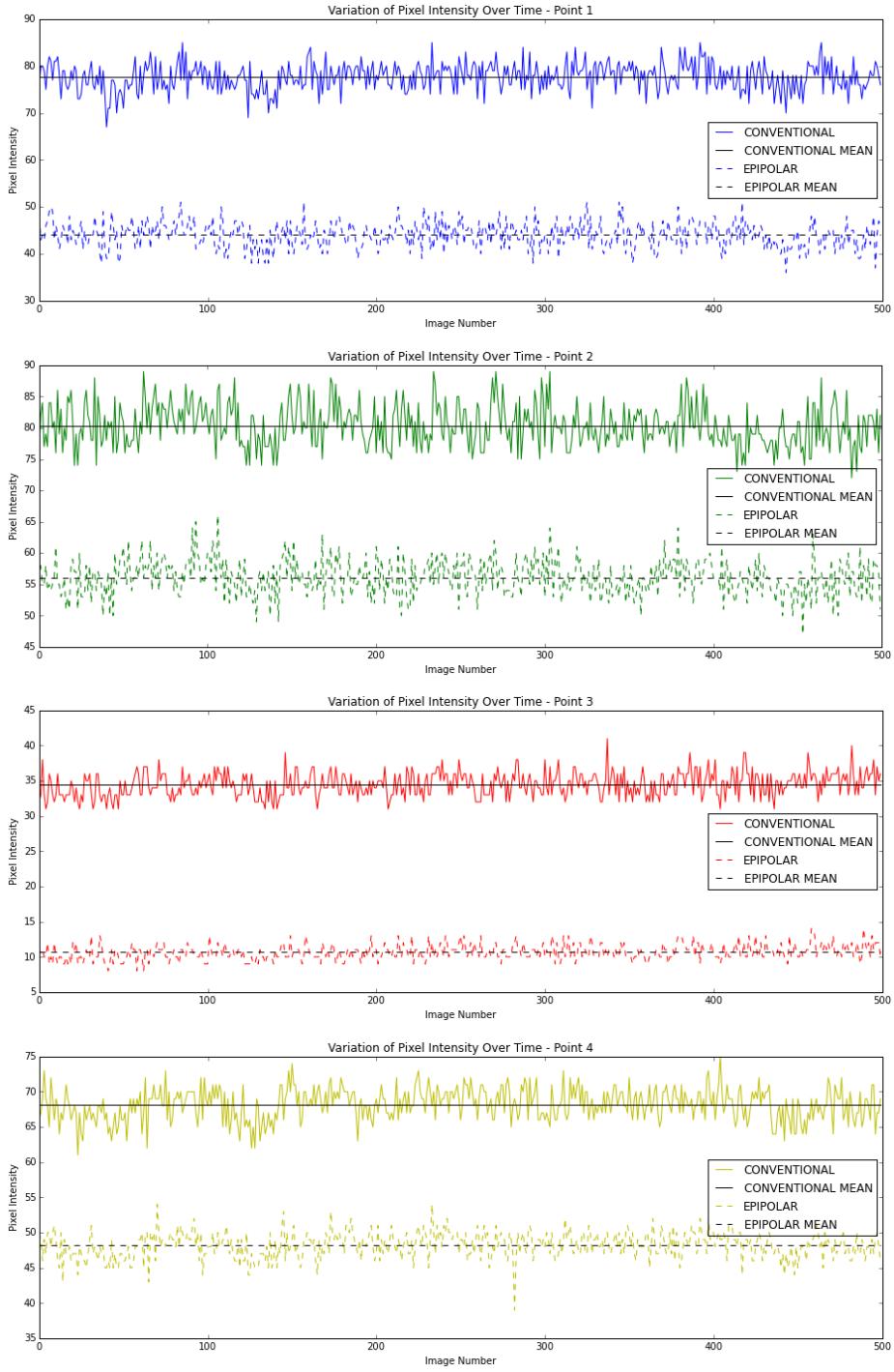


Figure 25. The intensity of each labeled pixel in Figure 24 is shown for all 500 images. The hardware limitations of the imaging system causes high variations in the recorded intensities of each pixel. Solid: Conventional Imaging. Dashed: Epipolar-Only Imaging. Black: The mean intensity over all images. Top to Bottom: The colour of the plots correspond to the labels in Figure 24.

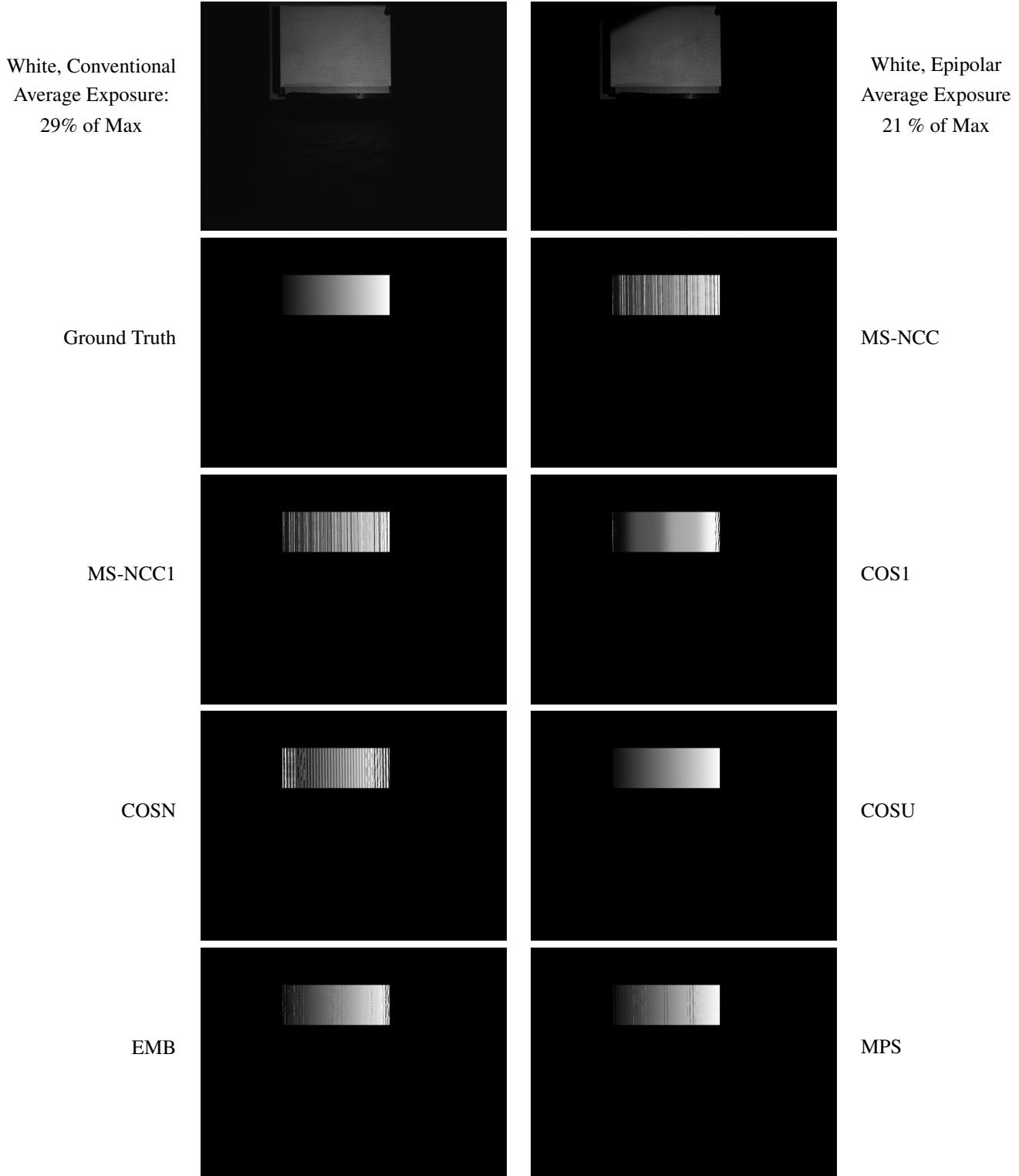


Figure 26. Top: A conventional image and an epipolar-only image of a flat board under white light. Due to a lack of reflections, there is not much of a difference between the two images. Since the imaging system does not perfectly capture the epipolar image, the effective imaging area is reduced and some light is blocked. This causes the dark band at the top and a lower exposure for the epipolar image. Middle to Bottom: The ground truth and the results of code matching from all methods are shown. The matching is done for 7 patterns, a maximum frequency of 40 (except COSU that is always the unit frequency), epipolar-only imaging, and each method's specialized matching algorithm is used (except for EMB where we use MS-NCC matching as the original method is not good for 7 patterns). Please zoom in for details.

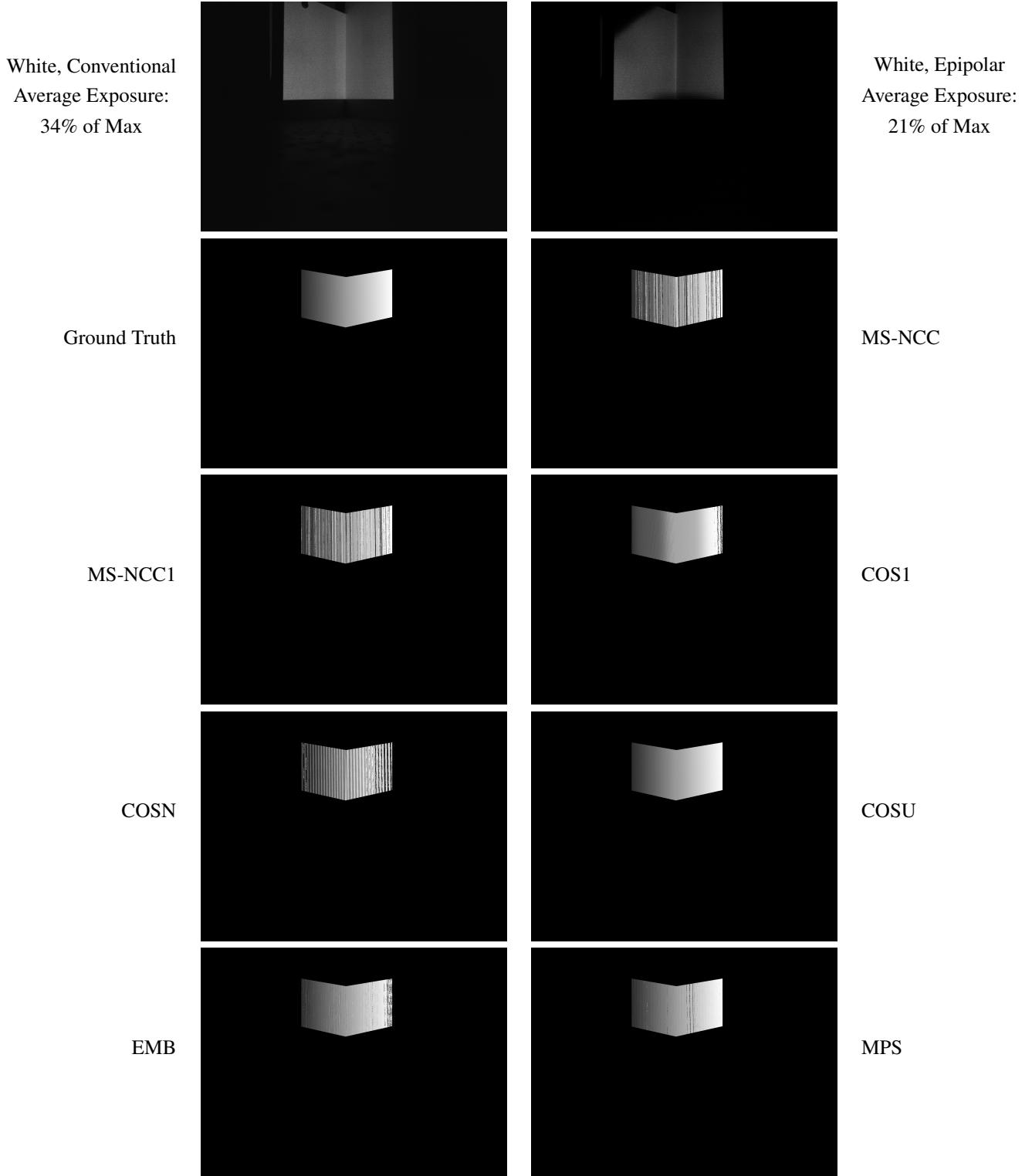


Figure 27. Top: A conventional image and an epipolar-only image of a wedge under white light. Notice how the area around the edge becomes darker in the epipolar-only image as the amount of interreflections increase, but are not captured. The edge itself is still bright as the light directly reflects off of it. Middle to Bottom: The ground truth and the results of code matching from all methods are shown. The matching is done for 7 patterns, a maximum frequency of 40 (except COSU that is always the unit frequency), epipolar-only imaging, and each method's specialized matching algorithm is used (except for EMB where we use MS-NCC matching as the original method is not good for 7 patterns). Please zoom in for details.

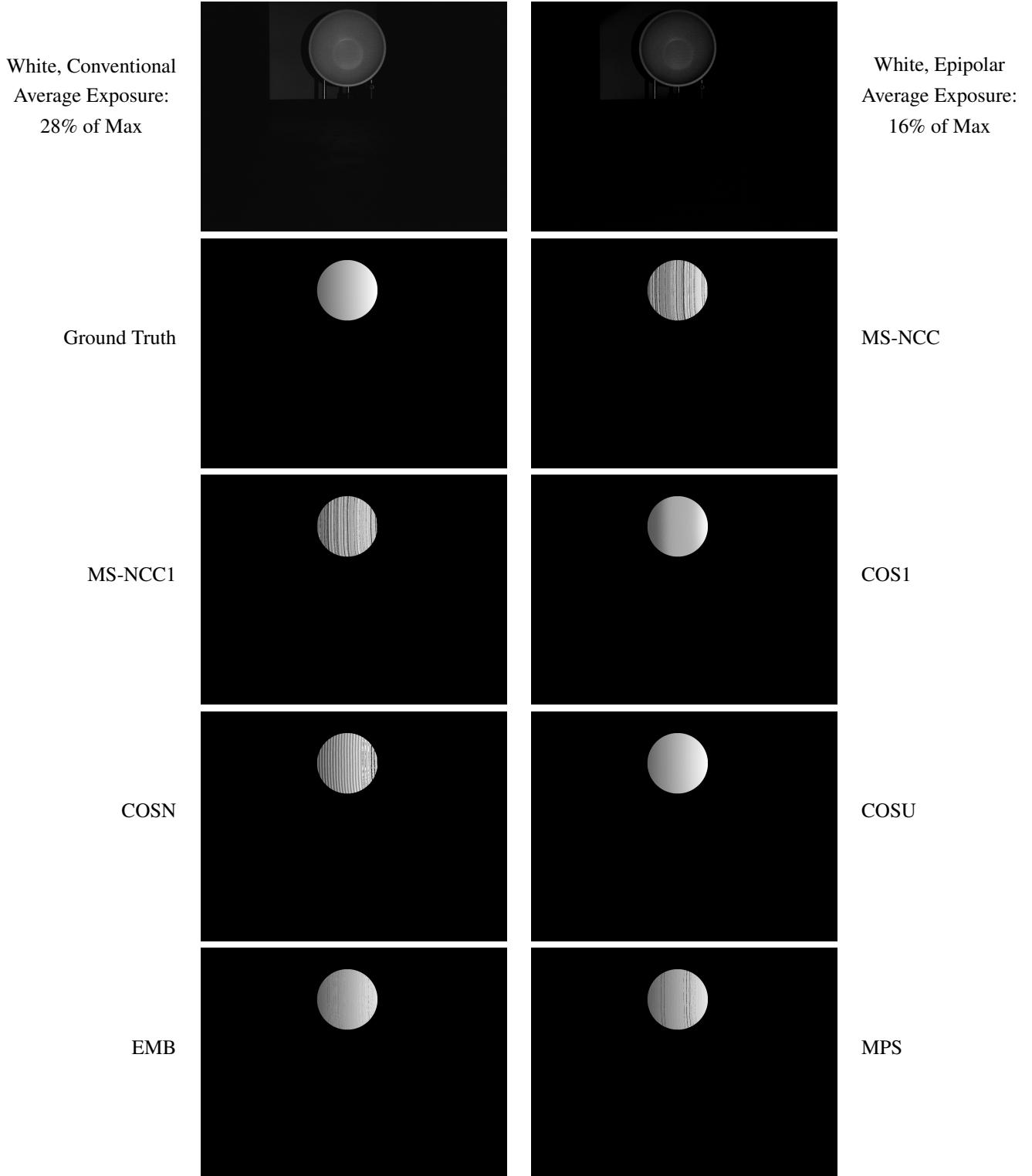


Figure 28. Top: A conventional image and an epipolar-only image of a bowl under white light. Notice how the whole bowl is brighter in the conventional image from the high occurrence of interreflections, especially near the rim where the surface is no longer close to parallel to the projector. Middle to Bottom: The ground truth and the results of code matching from all methods are shown. The matching is done for 7 patterns, a maximum frequency of 40 (except COSU that is always the unit frequency), epipolar-only imaging, and each method's specialized matching algorithm is used (except for EMB where we use MS-NCC matching as the original method is not good for 7 patterns). Please zoom in for details.

Flat Board, Epipolar-Only Imaging, Percentage of Incorrect Pixels

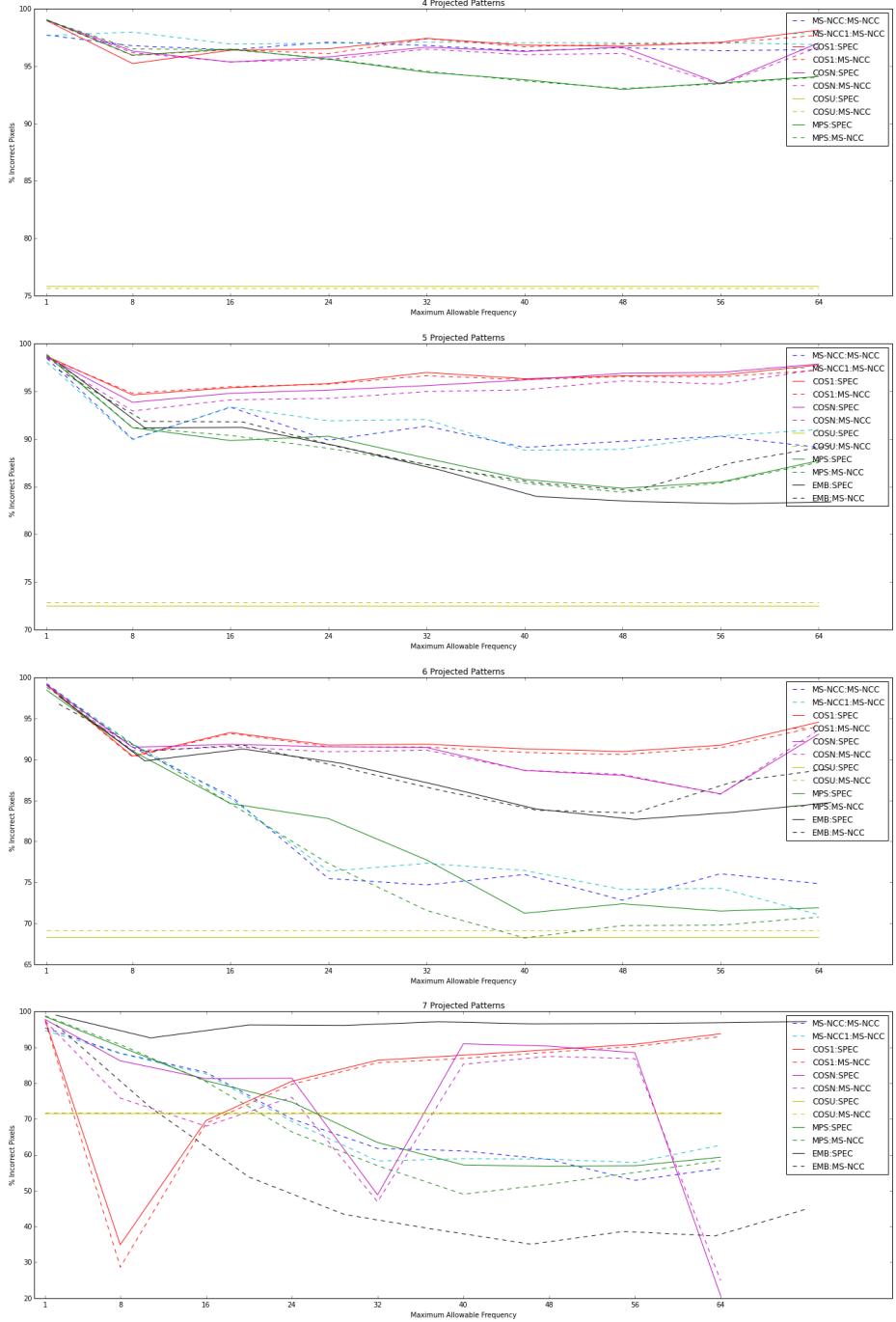


Figure 29. The percentage of incorrect pixels for the flat board captured with epipolar-only imaging. A solid line indicates the matching was done with the pattern's specialized matching method, while a dashed line indicates that the MS-NCC method was applied. The errors are similar regardless of the matching method apart from MPS, which benefits from MS-NCC, and EMB, which does a little worse with MS-NCC (except for 7 patterns). The projector blur causes high frequencies to do poorly, while the unit cosine is improved by virtue of not being affected. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively.

Flat Board, Epipolar-Only Imaging, RMS

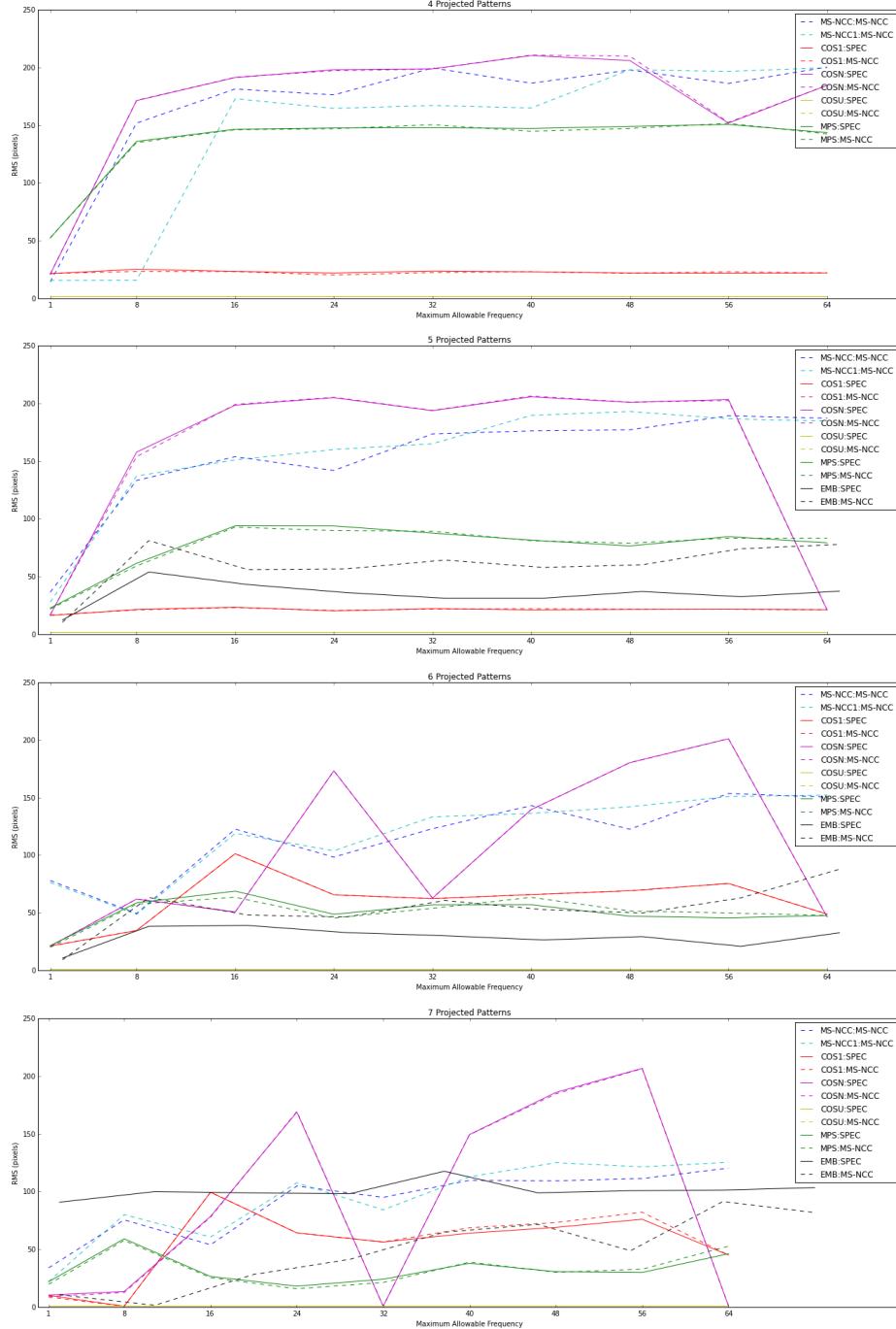


Figure 30. The RMS (in pixels) for the flat board captured with epipolar-only imaging. Solid: Specialized matching method. Dashed: MS-NCC matching. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The full range of errors are shown. See Figure 31 for a zoomed in version to distinguish between the methods that have the lowest error. The matching methods are the same apart from EMB. The unit cosine consistently has a low RMS and is only surpassed by other methods with 7 patterns.

Flat Board, Epipolar-Only Imaging, RMS, Zoomed

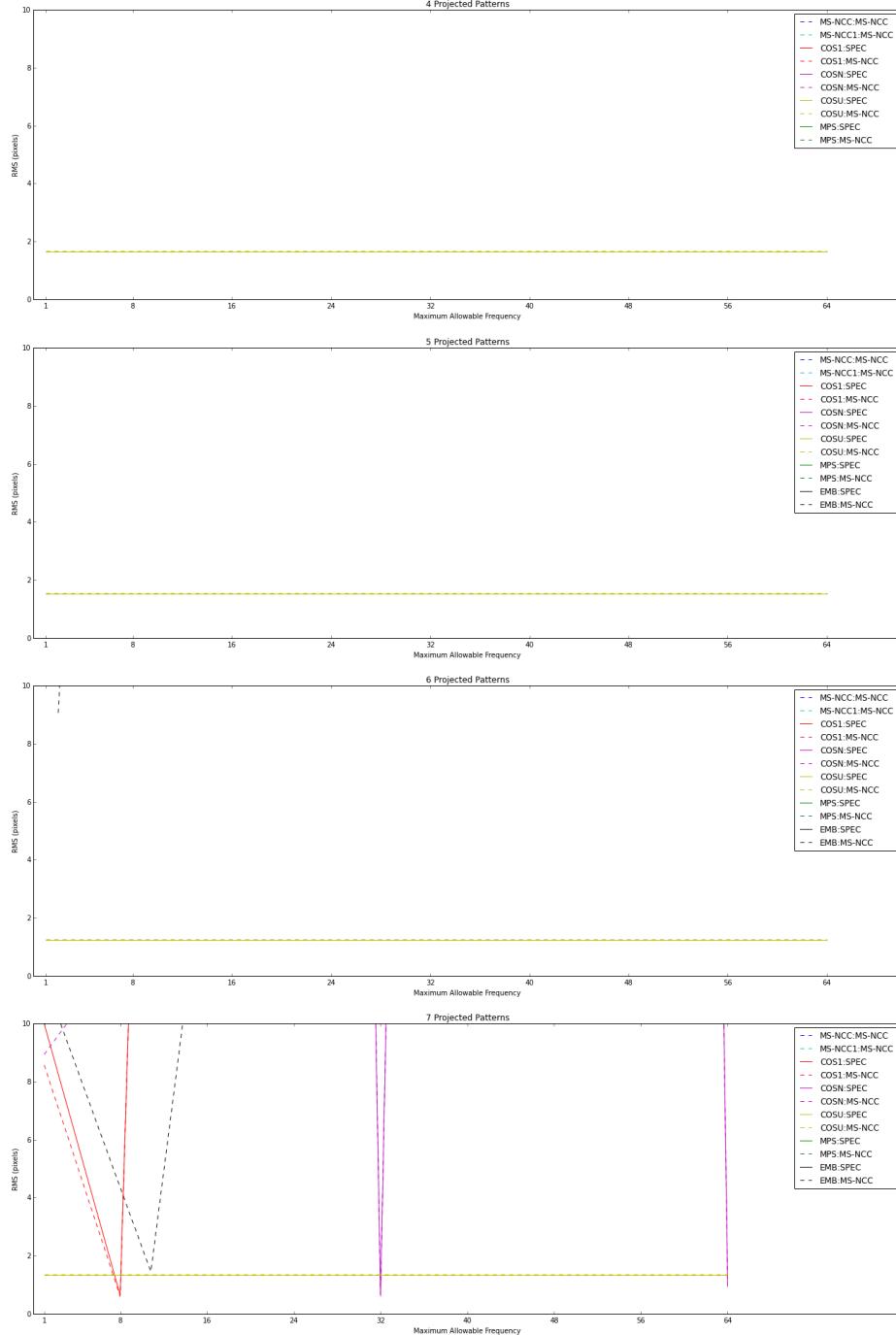


Figure 31. The zoomed in version (maximum RMS displayed is 10 pixels) of Figure 30 to distinguish between the methods that have the lowest error.

Wedge, Conventional vs Epipolar-Only Imaging, Percentage of Incorrect Pixels

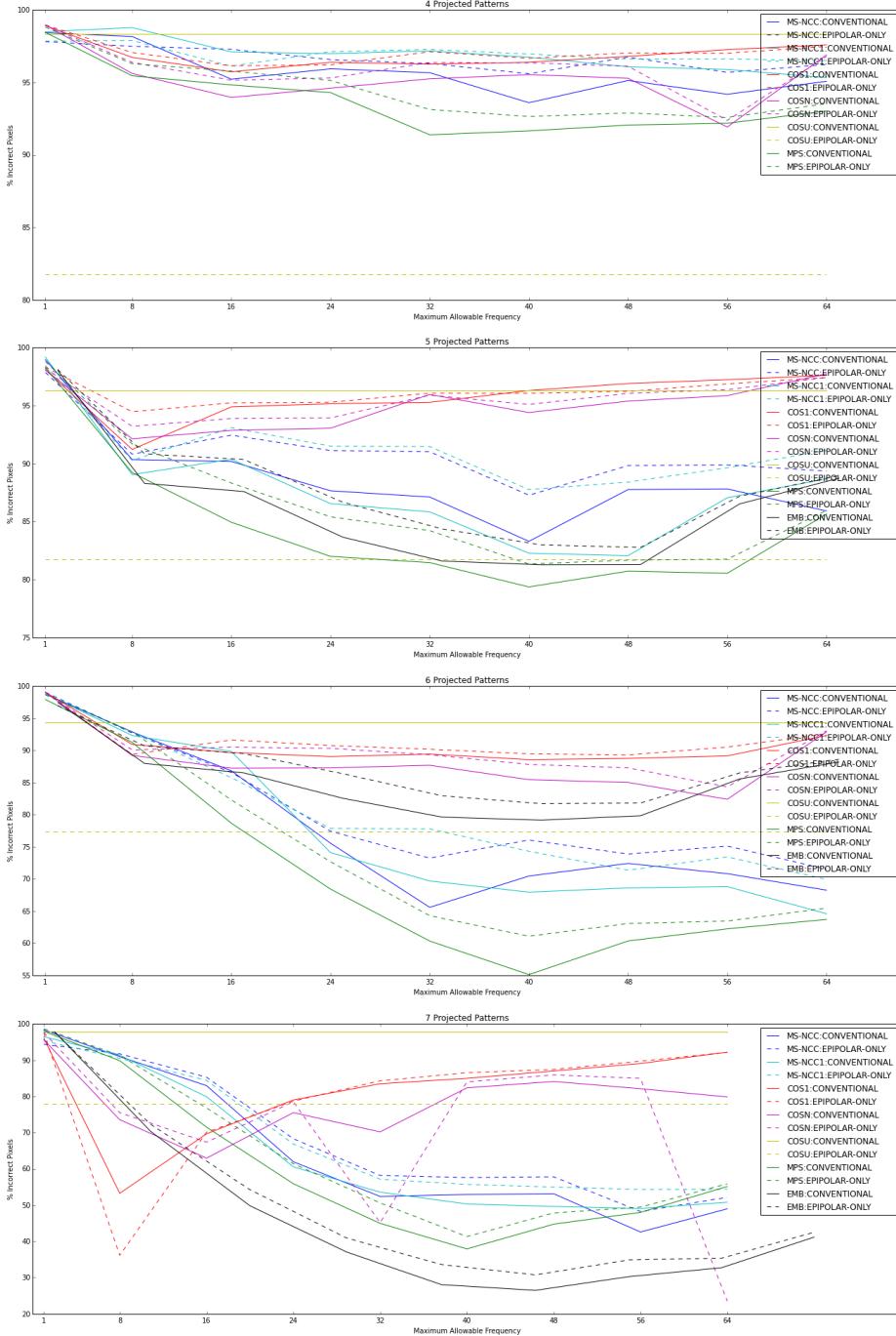


Figure 32. The percentage of incorrect pixels for the wedge matched using MS-NCC. A solid line indicates that the images were conventional, while a dashed line indicates that the images were epipolar-only. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. For most frequencies, conventional imaging has lower errors due to the hardware of the imaging system. For patterns that use the unit frequency to good effect, such as COSU, blocking the low frequency interreflections makes a significant difference.

Wedge, Conventional vs Epipolar-Only Imaging, RMS

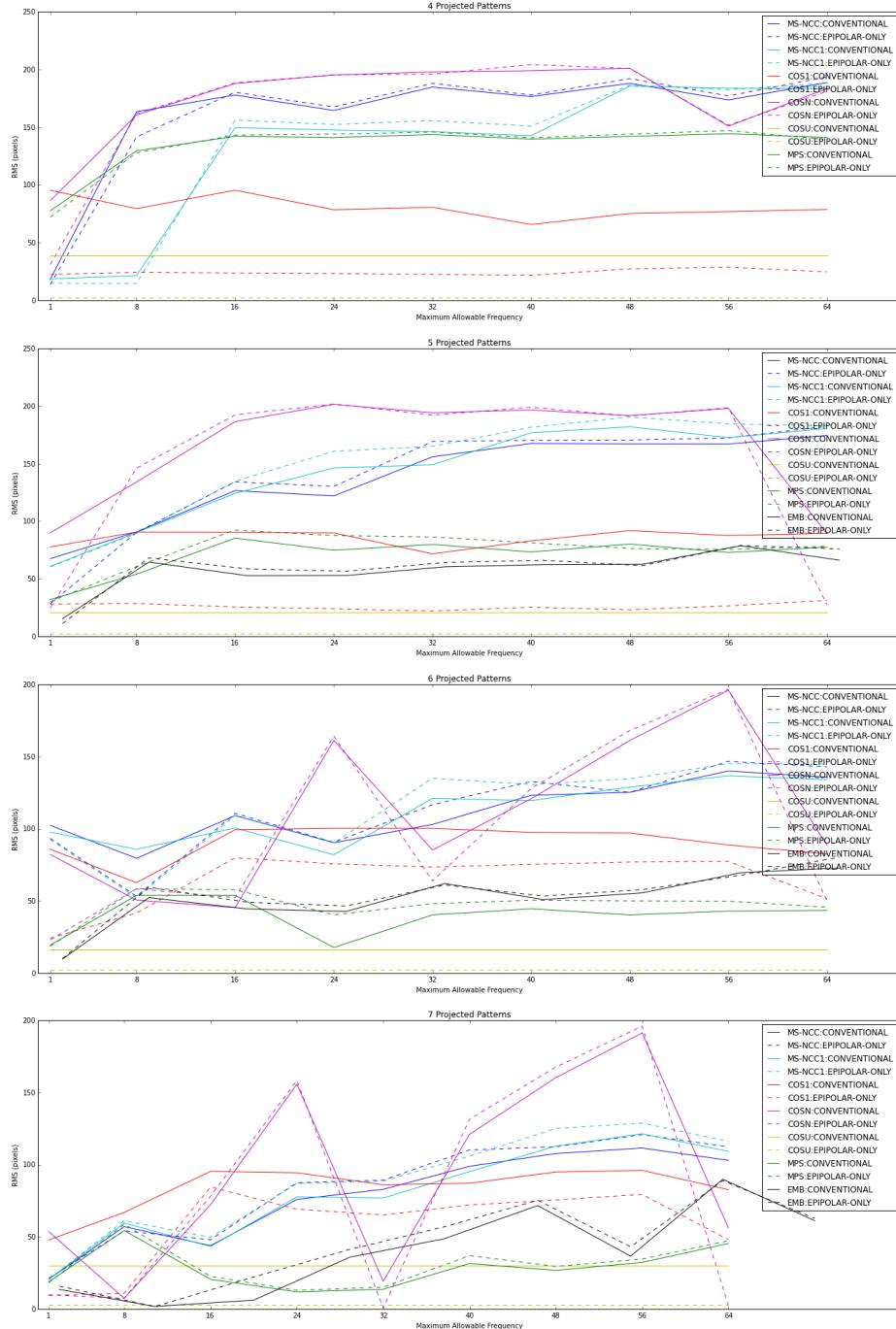


Figure 33. The RMS (in pixels) for the wedge matched using MS-NCC. A solid line indicates that the images were conventional, while a dashed line indicates that the images were epipolar-only. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The full range of errors are shown. See Figure 34 for a zoomed in version to distinguish between the methods that have the lowest error. It is clear in the RMS that patterns that use the unit frequency cosine greatly benefit from stopping interreflections. When that is done, the RMS is similar to that of a flat plane.

Wedge, Conventional vs Epipolar-Only Imaging, RMS, Zoomed

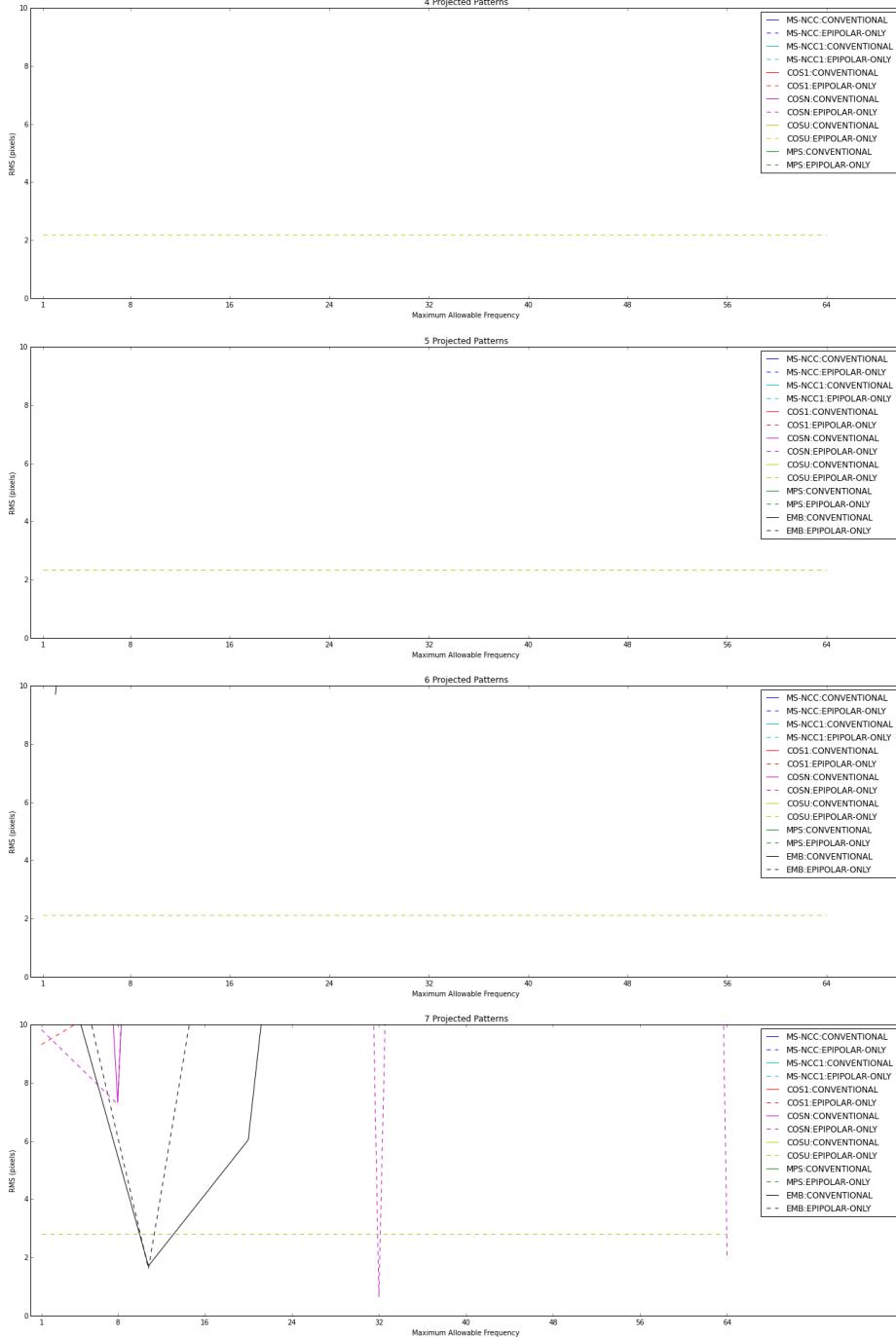


Figure 34. The zoomed in version (maximum RMS displayed is 10 pixels) of Figure 33 to distinguish between the methods that have the lowest error.

Bowl, Conventional vs Epipolar-Only Imaging, Percentage of Incorrect Pixels

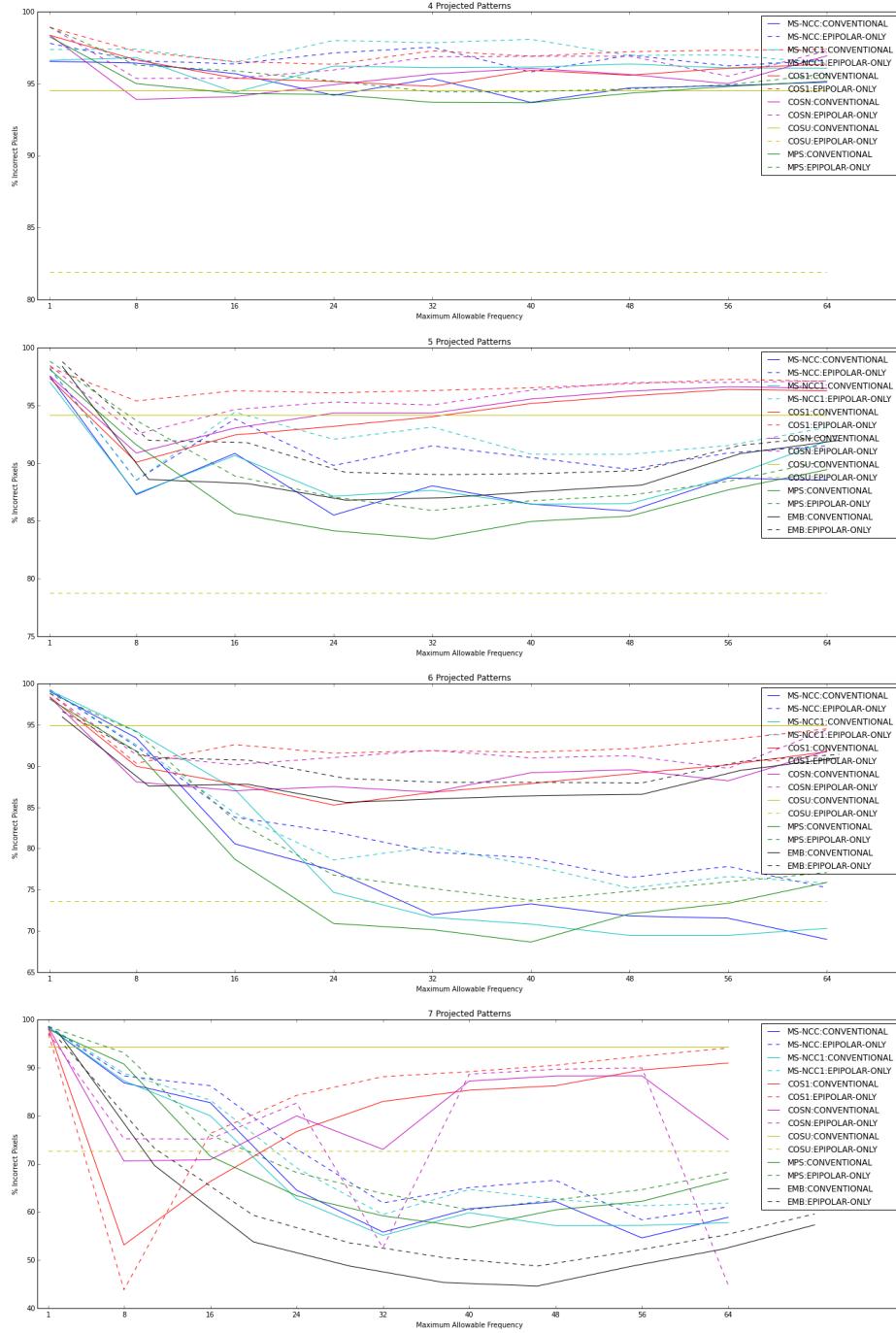


Figure 35. The percentage of incorrect pixels for the bowl matched using MS-NCC. A solid line indicates that the images were conventional, while a dashed line indicates that the images were epipolar-only. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The result is similar to the wedge, as shown in Figure 32.

Bowl, Conventional vs Epipolar-Only Imaging, RMS

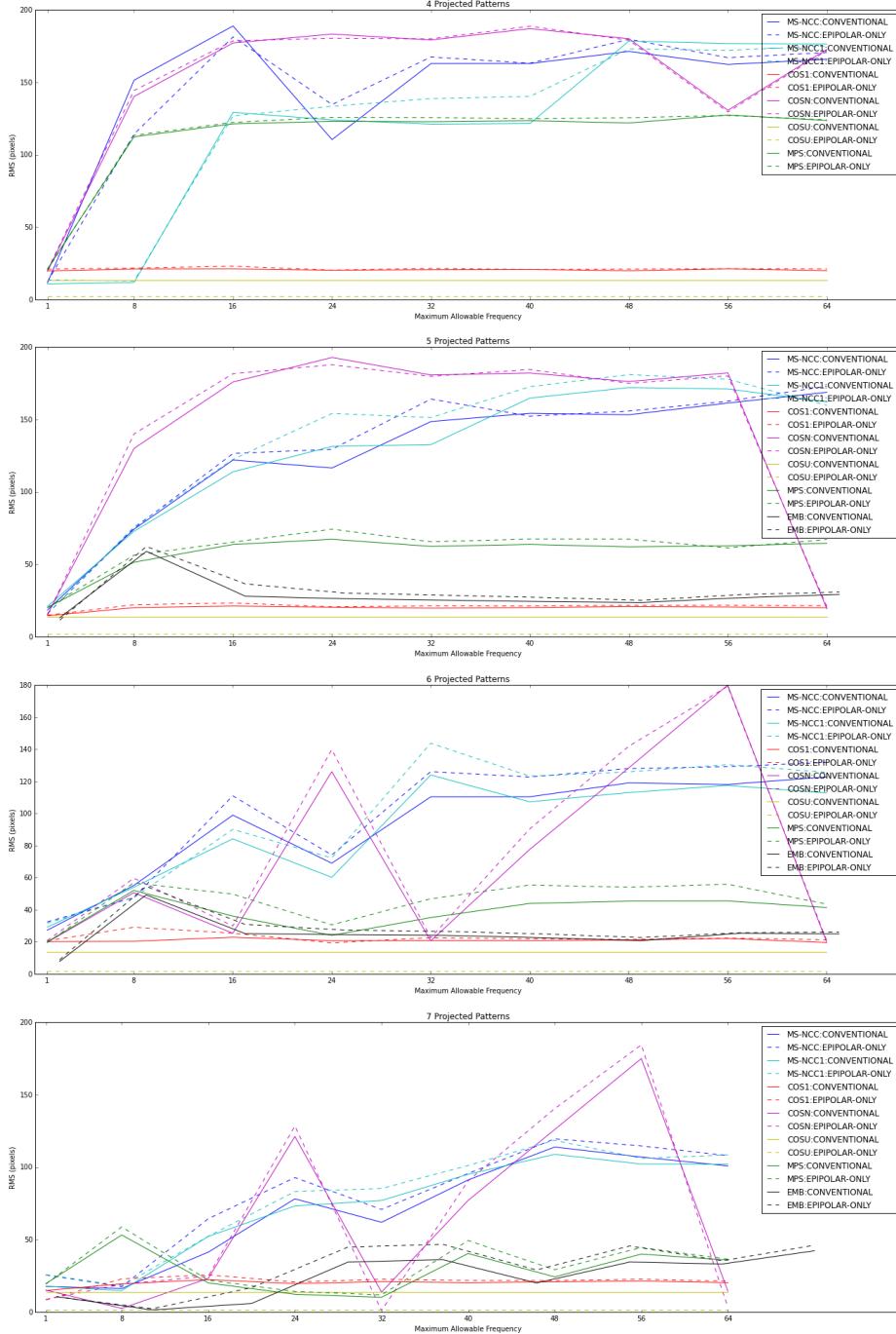


Figure 36. The RMS (in pixels) for the bowl matched using MS-NCC. A solid line indicates that the images were conventional, while a dashed line indicates that the images were epipolar-only. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The full range of errors are shown. See Figure 37 for a zoomed in version to distinguish between the methods that have the lowest error. The result is similar to the wedge, as shown in Figure 33.

Bowl, Conventional vs Epipolar-Only Imaging, RMS, Zoomed

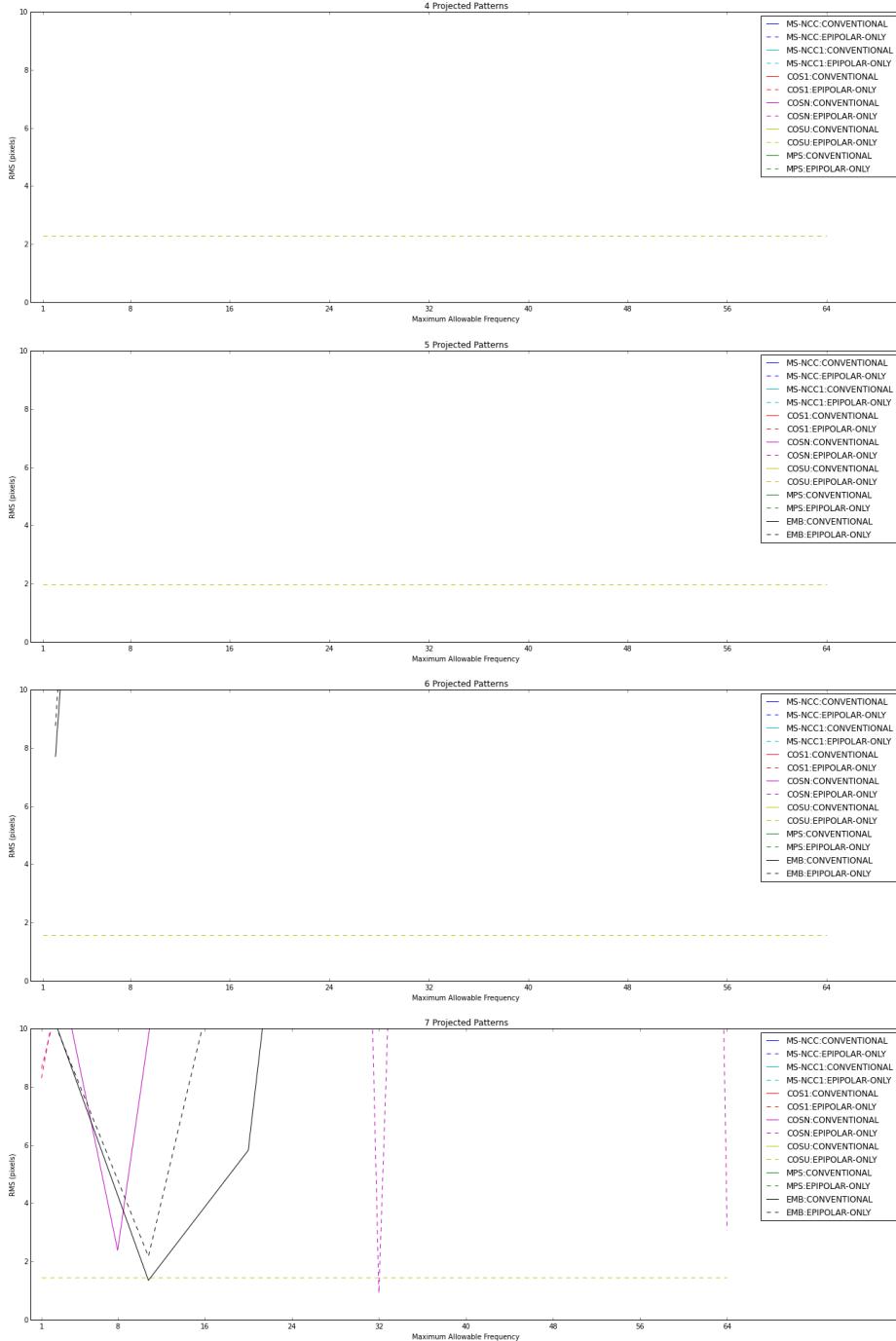


Figure 37. The zoomed in version (maximum RMS displayed is 10 pixels) of Figure 36 to distinguish between the methods that have the lowest error.

Wedge vs Bowl, Conventional Imaging, Percentage of Incorrect Pixels

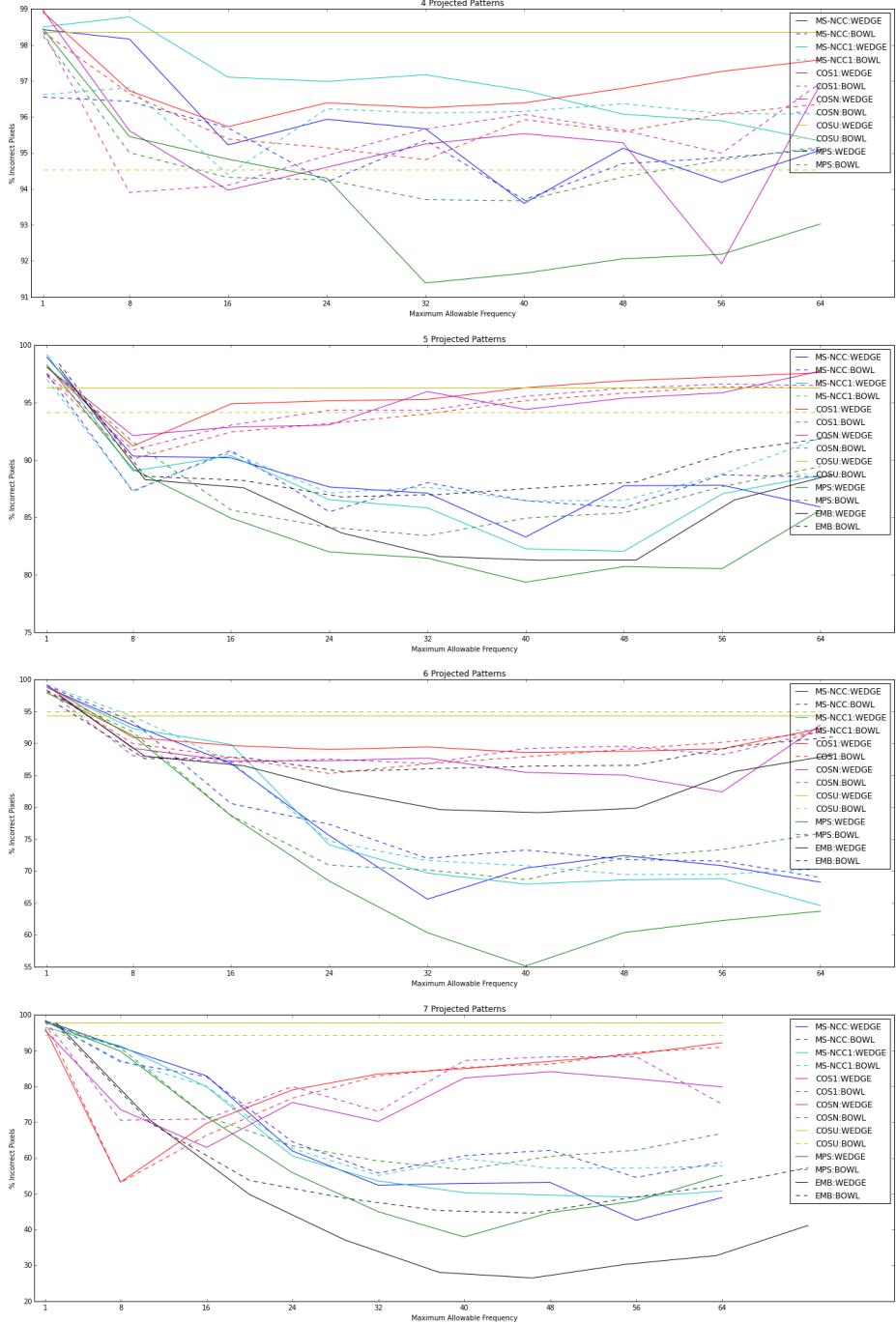


Figure 38. The percentage of incorrect pixels for a conventionally taken image matched using MS-NCC. A solid line indicates that the images were from the wedge, while a dashed line indicates that the images were from the bowl. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The wedge generally has lower errors than the bowl. Since the errors are localized to an edge, there will be a lower occurrence of errors for the wedge than for the bowl.

Wedge vs Bowl, Conventional Imaging, RMS

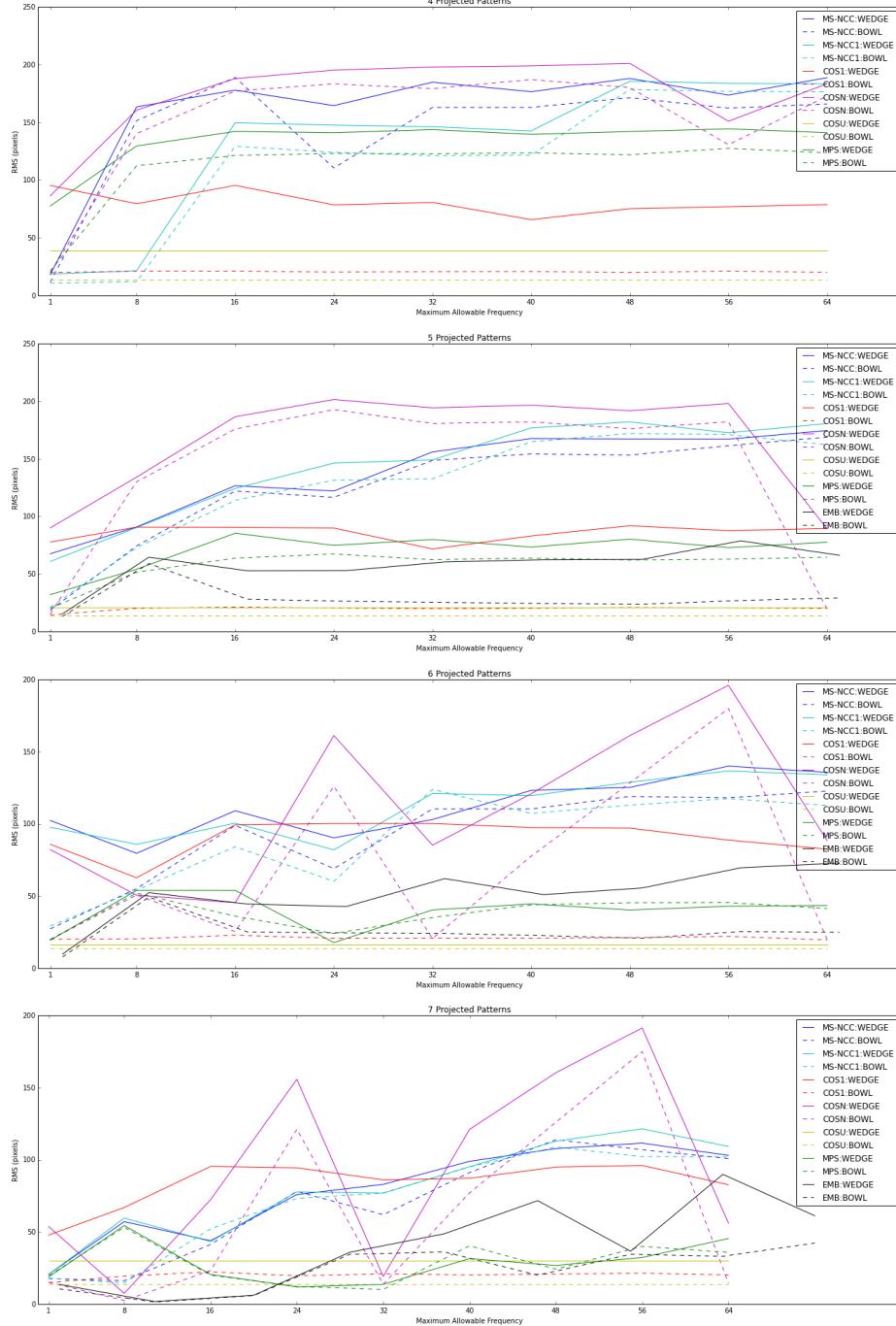


Figure 39. The RMS (in pixels) for a conventionally taken image matched using MS-NCC. A solid line indicates that the images were from the wedge, while a dashed line indicates that the images were from the bowl. Top to Bottom: The number of patterns used are 4, 5, 6, and 7, respectively. The full range of errors are shown. See Figure 40 for a zoomed in version to distinguish between the methods that have the lowest error. The bowl has lower errors than the wedge. Due to the stronger interreflections near the edge, pixels beside it may have their labels changed greatly. Even with a higher occurrence of errors, the bowl has a better RMS due to its weaker interreflections.

Wedge vs Bowl, Conventional Imaging, RMS, Zoomed

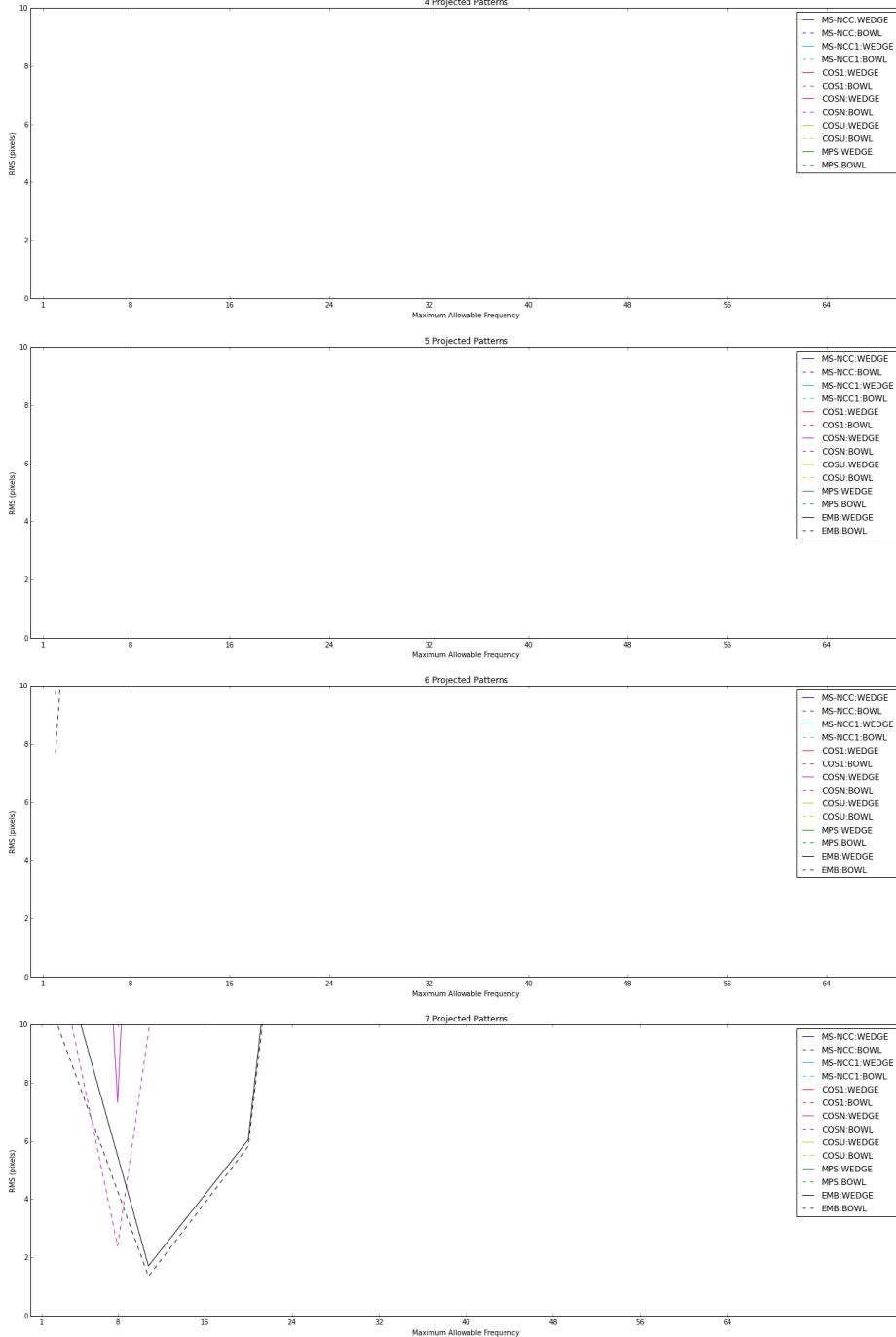


Figure 40. The zoomed in version (maximum RMS displayed is 10 pixels) of Figure 39 to distinguish between the methods that have the lowest error.