Global Illumination-Invariant Fast Sub-Pixel Image Registration

Andrew Gilman
Institute of Natural and Mathematical Sciences
Massey University
Auckland, New Zealand
Email: a.gilman@massey.ac.nz

Arno Leist
School of Engineering and Advanced Technology
Massey University
Auckland, New Zealand
Email: a.leist@massey.ac.nz

***Abstract*—Sub-pixel image registration is an important part of many image processing and computer vision applications. We propose a computationally simple direct (i.e., non-iterative) method for sub-pixel registration of translated images. To register two images, first a global space-invariant resampling filter is computed that is a least-squares optimal predictor of one image from the pixel values of the other image. Then, the coefficients of this filter are linearly combined to compute the offset between the two images. The computational cost of this algorithm is linear in the number of pixels. The accuracy and efficiency of the proposed algorithm is demonstrated to be better than a range of existing methods for images with various levels of high-frequency detail and at various noise levels.**

***Keywords*—image registration; sub-pixel; direct; least-squares optimal; linear computational complexity.**

I. INTRODUCTION

Image processing often requires processing of data from multiple images of the same scene captured from different viewpoints with one or multiple cameras or maybe even with completely different types of sensors. Before this can be achieved, the pixel data in the images must be somehow 'synchronised' and this is achieved using registration. Image registration is the problem of finding a geometric transformation that maps the coordinate plane of one image to another using the image data itself. It is an important problem in image processing and computer vision, required as part of many applications in areas such as medical imaging, remote sensing and consumer electronics [1].

What makes this problem challenging is the fact that many applications require fast, yet accurate, registration. For some applications, such as image super-resolution or image fusion (see Figure 1 for an example), it is critical to register input images with accuracy down to sub-pixel level. But, unfortunately, speed or computational complexity and accuracy are generally trade-offs and can be hard to achieve simultaneously. Sub-pixel registration can also be hampered by the fact that the images may be contaminated by noise and aliasing, both of which may have a considerable effect, as well as possible lack of conformity of the image data to the proposed transformation model.

Even for applications that do not require precise sub-pixel registration, it is still very important, because registration of images with larger integer offsets is often achieved using a multi-resolution image pyramid [2] for efficiency reasons (for examples see [3]–[5]). Ignoring the high-resolution detail at higher levels of the pyramid is not only more computationally efficient, but also necessary, because of aliasing of high spatial frequency components undergoing large motion. The use of a multi-resolution framework helps to alleviate this problem [6].

The typical approach to solving the registration problem is to place it into an optimisation framework [7], with the objective function defined by some metric that measures similarity (or alternatively distance) between one image and a transformed version of the other. The objective function is then optimised with respect to the transformation parameters. This approach was popularised by Lucas and Kanade [8] in 1981 and has seen many variants proposed since, such as using different optimisation methods and similarity metrics. A summary of some common similarity metrics and search strategies can be found in [1] and [9]. While many of these approaches are theoretically sound, they are generally computationally expensive and require a number of iterations to achieve convergence, due to the non-linear nature of the problem. This is highly inadequate for some applications that require the results to be computed in a timely manner. An example of such applications, which is becoming increasingly popular, is image processing (High Dynamic Range (HDR) [10], panorama stitching, etc.) on smartphone and tablet devices that posses only limited computing resources.

We are interested in a special case of image registration, where the geometric transformation is constrained to pure translation. The desire for fast registration of translated images with sub-pixel accuracy has led to the development of a number of direct approaches that attain lower complexity

Fig. 1: High Dynamic Range image produced from registered images with varying exposures.

 ISBN: 978-1-61208-283-7

(a) face (b) cat (c) bird (d) beach

Fig. 2: Test images

by making various simplifying assumptions. Fourier phase correlation [11]–[14], for example, assumes that the images obey the Nyquist criterion [15], while interpolation of the cross-correlation function [16] makes assumptions about the shape of the scene's autocorrelation function. Unfortunately, when the real data deviates from the assumptions made in the model the accuracy can suffer.

We present a new direct technique for sub-pixel registration of translated images. Even though this method does not require iteration, its formulation has more similarities with the iterative methods than the direct, although it does not require the choice of an appropriate interpolation function for the resampling filter. Instead, it computes the optimal resampling filter from the image data itself using linear least-squares. The relative shift between the images is then computed in a novel way from the filter coefficients.

In Section II, we introduce the theory of image registration in the optimisation framework and present the new direct approach. In Section III, we describe the experimental method for sub-pixel accuracy evaluation. Section IV presents the experimental results and their discussion, followed by conclusions in Section V.

II. Theory

In this section, we provide a mathematical definition for the image registration problem and show how it is formulated in the optimisation framework, which requires a computationally intensive iterative solution. We then consider an alternative formulation, which results in a direct solution.

Let us consider the problem of registering a discrete reference image $f(kT, lT)$ and an offset target image $g(kT, lT) = f(kT + u_x, lT + u_y)$, where k and l are integer pixel indices, T is the pixel pitch (which we will assume to be equal to one) and u_x, u_y are some unknown and possibly non-integer offsets between the two images. The goal of sub-pixel image registration is to estimate u_x and u_y with as much accuracy as possible. A conventional way of doing this is to iteratively minimise the difference (usually sum of squared differences) between $f(k + \hat{u}_x, l + \hat{u}_y)$ and $g(k, l)$:

$$\arg\min_{\hat{u}_x, \hat{u}_y} \sum_{k,l} \left(g(k,l) - f(k + \hat{u}_x, l + \hat{u}_y)\right)^2 \quad (1)$$

In practice, since the image $f(k, l)$ is only known on an integer grid, some sort of interpolation must be applied to $f(k, l)$ to get the shifted image $f(k + \hat{u}_x, l + \hat{u}_y)$. For computational reasons, only the values located on the shifted grid are interpolated and this procedure is generally performed using a resampling filter [17]. The filtering operation can be expressed as

$$\sum_{m=M^-}^{M^+} \sum_{n=N^-}^{N^+} f(k + m, l + n)\, h_{\hat{u}_x, \hat{u}_y}(m, n) \quad (2)$$

where $m, n \in \mathbb{Z}$ and $h_{\hat{u}_x, \hat{u}_y}(m, n)$ is a resampling filter, based on some predetermined interpolation basis function (see [18], [19] for some common examples), with a window span between M^- and M^+ in x direction and between N^- and N^+ in y direction. The resampling filter coefficients are samples of the interpolation basis function offset by (u_x, u_y) and are generally non-linear in u_x and u_y, making expression (1) a non-linear least-squares problem. Choosing a different interpolation basis function would result in a slightly different interpolated image, in turn affecting the results of minimisation.

A. Registration through Optimal Interpolation

Now, let us consider an alternative formulation of this problem. First, reformulate (1) to minimise with respect to the filter coefficients $h(m, n)$ instead of the offsets u_x and u_y. This removes any dependence on an interpolation basis function and results in filter coefficients being optimised in a least-squares fashion. The resulting filter is optimal – no other resampling filter can do a better job of predicting image $g(kT, lT)$ from $f(kT, lT)$.

$$\arg\min_{\mathbf{h}} \sum_{k,l} \left(g(k,l) - \sum_m \sum_n f(k + m, l + n) h(m, n)\right)^2 \quad (3)$$

The next step is to determine the relative offsets u_x and u_y. Given a set of filter coefficients, computed from an interpolation basis function, we could compute the offset by which this particular filter would shift the image through a linear combination of the filter coefficients. Consider the Keys' cubic convolution interpolation kernel:

$$CC(x) = \begin{cases} \frac{3}{2}|x|^3 - \frac{5}{2}|x|^2 + 1 & 0 < |x| < 1 \\ -\frac{1}{2}|x|^3 + \frac{5}{2}|x|^2 - 4|x| + 2 & 1 < |x| < 2 \\ 0 & 2 < |x| \end{cases} \quad (4)$$

(a) Image 'bird'. Noise std: 10.8, 28.8, 46.8

(b) Image 'face'. Noise std: 4.6, 12.3, 20.0

Fig. 3: Effect of noise strength on test images. Noise std stated in grayscale levels.

and a one-dimensional resampling filter $h_u(m)$, based on this kernel, defined as

$$h_u(m) = CC(-u + m), \quad m \in \mathbb{Z}. \tag{5}$$

It can be shown algebraically (see Appendix) that the offset u can be recovered from the filter coefficients through a weighted sum combination thereof:

$$u = \sum m h_u(m) \tag{6}$$

This is also true for the optimal filter coefficients, which do not depend on any explicit interpolation function, but are still related to the offset between the images through the image data itself. The offset estimate between the images can be computed in the same way as:

$$\hat{u}_x = \sum_{m=M^-}^{M^+} \sum_{n=N^-}^{N^+} m h(m,n) \tag{7}$$

$$\hat{u}_y = \sum_{m=M^-}^{M^+} \sum_{n=N^-}^{N^+} n h(m,n) \tag{8}$$

and we show that this works experimentally, in Section IV.

B. Global Illumination Invariance

Cameras with automatic exposure setting will meter the scene prior to taking a picture. Capturing multiple images of the scene, even in quick succession, can result in global illumination variations due to the metering algorithm calculating slightly different exposure settings. This difference can be fairly large in some cases – for example when capturing different parts of the scene, such as when panning the camera to capture a panorama. Using different exposure settings may also be intended, for example, when capturing images to produce a high-dynamic range reproduction of a scene. The image processing algorithms, including registration, must be able to cope well with these situations.

Global illumination difference between the two images under registration can be accounted for in the model by including gain A and offset S: $g(kT, lT) = Af(kT + u_x, lT + u_y) + S$. Because the filter coefficients in optimisation expression (1) are not constrained to sum to one, this optimisation would automatically account for gain A, resulting in

$$\sum_m \sum_n h(m,n) = 1/A \tag{9}$$

The relative offsets u_x and u_y can still be recovered from $h(m,n)$ by normalising expressions (7) and (8) by the sum of filter coefficients:

$$\hat{u}_x = \frac{\sum_m \sum_n m h(m,n)}{\sum_m \sum_n h(m,n)} \tag{10}$$

$$\hat{u}_y = \frac{\sum_m \sum_n n h(m,n)}{\sum_m \sum_n h(m,n)} \tag{11}$$

The illumination offset S can be accounted for by including an additional variable $\hat{S}$ into the optimisation procedure (1), subtracted from $g(k,l)$. This does not change the equations (10) and (11), by which the offset is estimated from the filter coefficients.

III. Method

To validate the proposed method and assess its sub-pixel accuracy, we have performed an experiment using synthetically

 ISBN: 978-1-61208-283-7

generated images with known sub-pixel offsets.

A. Test Image Generation

A high-resolution image (10 times higher resolution than the test images) was taken as the source scene. A number of test scenes (shown in Figure 2) that contain different combinations of texture, sharp detail and smooth areas at a range of different scales were selected for this performance evaluation. These scenes vary in high-frequency content, resulting in different amounts of aliasing in the test images, from little aliasing in image 'beach' to a lot of aliasing in image 'bird'.

A simple model of the imaging process was used to simulate the capture of lower resolution test images. The imaging model consisted of area-sampling (any blurring due to the optics' Point-Spread Function (PSF) was assumed to be insignificant in comparison to the degradation from the sensor PSF – this assumption is not unrealistic for many consumer cameras) and additive white Gaussian noise [20]. Creation of test images in this way allowed for aliasing to occur providing the opportunity to test the algorithms on aliased data, rather than just with band-limited signals.

The 1300×1300 pixel grayscale source images with 8-bit precision were filtered using a 10×10 box average to simulate blurring during area-sampling, and then down-sampled by a factor of 10 in each dimension to result in 124×124 pixel test images (low-resolution images were cropped to avoid edge effects from the filtering operation). Shifting the high-resolution image by an integer number of pixels prior to downsampling allowed for generation of low-resolution test images shifted by fractional (in steps of 0.1) amounts of a pixel.

Here, we test the ability of the method to register offsets smaller than one pixel by registering a set of one hundred image pairs with relative offsets ranging from $(0, 0)$ to $(0.9, 0.9)$.

B. Measurements

The uncertainty in estimating the relative offsets is not constant across the sub-pixel range. Registering image pairs that relate by a small sub-pixel translation (close to zero) or a translation of almost one whole pixel results in smaller MSE than registering a pair of images offset by about half a pixel (see Figure 4 for a typical error pin plot). Because there is no prior knowledge about the distribution of the translations, we estimate the performance of sub-pixel registration on average by assuming that all sub-pixel translations are equally likely to occur. Average performance can then be calculated simply by averaging the squared error over all possible sub-pixel translations – 100 translations, ranging from (0,0) to (0.9, 0.9) for this experimental set-up.

Each pair of images with a unique offset was registered 100 times, each time with a different instance of white Gaussian noise added to the images. A different pair of illumination gain A and offset S constants were also used each time to adjust the exposure of the target image. The gain constant was drawn from a normal distribution with a mean of one and a standard deviation of 0.1, and the offset constant was drawn from a normal distribution with a mean of 0 and a standard deviation of 25. The target image was clipped to only contain values between 0 and 255 after the transformation.

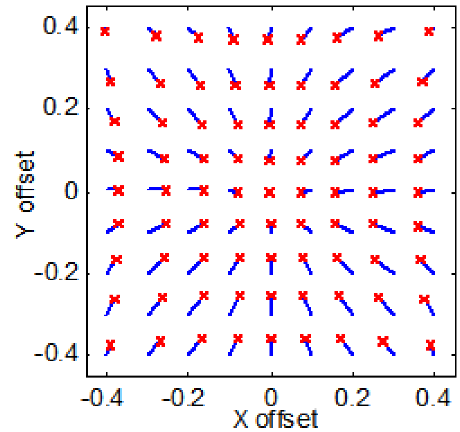

Fig. 4: Sub-pixel registration errors depicted as a pin plot for image 'bird' with no noise, using normalised cross-correlation method.

To make the results easier to interpret, registration error magnitude (12) was computed for each registration, and these were averaged over the 100 repeated results for each possible sub-pixel offset, and over all 100 sub-pixel offsets to give expected sub-pixel accuracy when registering two images.

$$||e||_2^2 = (u_x - \hat{u}_x)^2 + (u_y - \hat{u}_y)^2 \qquad (12)$$

The measurements were repeated at different noise levels. We found the sub-pixel accuracy of registration to be very sensitive to the noise level. To be able to compare the results between the test scenes, different amounts of noise were added in each case. This was achieved empirically by matching the visual effect of noise on each scene; see Figure 3 for an example – image 'bird' required almost twice the noise standard deviation in comparison to image 'face' to achieve the same level of contamination.

C. Methods for Comparison

To put the experimental results in perspective, we compare them to three other popular methods: one standard iterative method and two direct methods. We employ Lucas-Kanade iterative registration [8] with bilinear interpolation for resampling. A detailed description of implementation of this method can be found in [21].

The two direct techniques that are used are the normalised cross-correlation and a Fourier phase based method by Stone et al. [22]. The Fourier phase method was chosen because it works well in the presence of aliasing. Sub-pixel accuracy of normalised cross-correlation was achieved by interpolating a 3×3 window around the function peak using a 2D second order polynomial.

IV. Results and Discussion

We have performed the experiment with the proposed method using resampling filters of various sizes. We were able to successfully recover the relative shift between the images from the optimal filter coefficients according to equations 10 and 11, confirming our observations at the end of Section II-A.

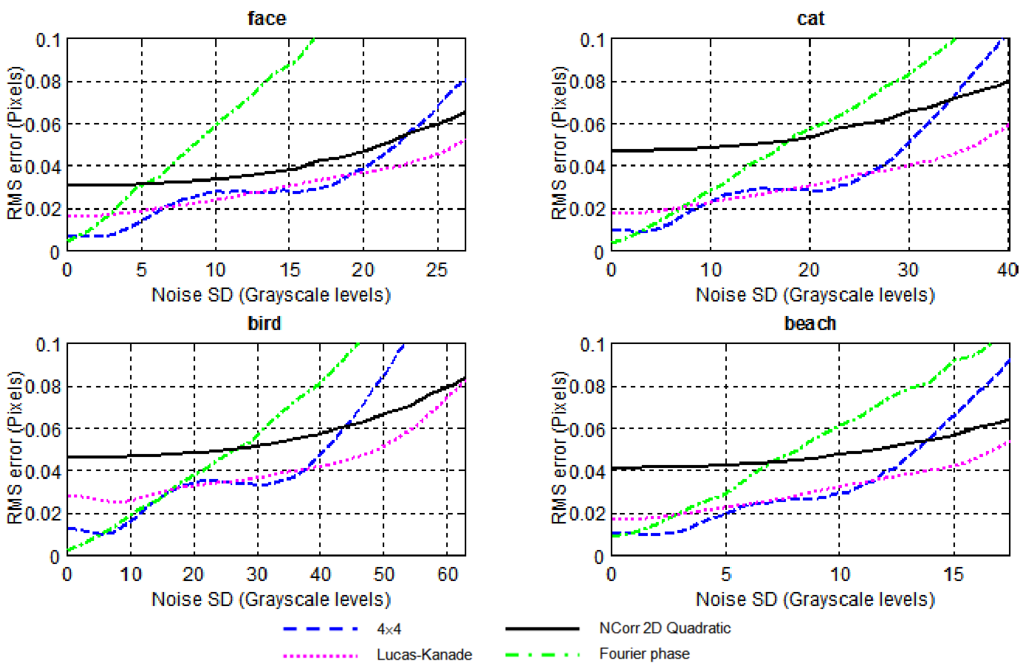

Fig. 5: Effect of noise strength on test images. Noise std stated in grayscale values.

Using filters of different sizes resulted in slightly different degrees of accuracy.

We report the results for the 4×4 filter, as it had the best performance relative to its computational complexity – larger filters obviously require more computation. A plot of root-mean-square registration error versus the noise standard deviation can be seen in Figure 5. The registration error is expressed as a fraction of the spatial pixel size (same as the pixel pitch), whereas the noise standard deviation is expressed in grayscale levels.

It can be observed that the proposed method can achieve an accuracy of as high as 1% of a pixel in a low-noise situation (left part of the graph). As the noise level increases, the accuracy drops slightly to around 4% of a pixel (middle part of the graph). At even higher noise levels, the accuracy drops off more quickly, but at these noise levels the images are quite heavily contaminated by noise, as can be seen from Figure 3.

Comparing the results of the proposed method to that of the other methods one can observe that it performs significantly better than the cross-correlation method, better than Lucas-Kanade method and better than the Fourier phase-based method, apart from images 'cat' and 'bird' with no added noise. These are favourable results for the proposed method, taking into consideration that it was twice as fast as normalised cross-correlation and Fourier phase-based methods and eight times faster than the iterative Lucas-Kanade method. All of these methods, except Fourier phase-based method, see a linear increase in execution time with an increase in image size. Fourier phase-based method is in theory $O(N \log N)$; however, because the transform is separable and image dimensions are small, it also sees almost linear increase in execution time with image size.

V. Conclusion

Image registration is traditionally an optimisation problem that makes use of interpolation as a crucial link between the discrete and continuous domains. The use of interpolation functions that are common in image processing, such as piece-wise cubic polynomials for example, leads to a non-linear objective function, the optimisation of which requires an iterative solution. In the case of sub-pixel registration of translated images, some direct methods also exist, but these make certain assumptions that can limit the accuracy if the assumptions do not hold. The optimal filter based image registration proposed in this paper is an alternative two step method for sub-pixel registration of translated images that does not require iteration. The first step computes a resampling filter that is a least-squares optimal predictor of the target image using the pixel values of the reference image. The second step estimates the translation between the target and the reference from the filter coefficients. The method has a linear-time computational complexity.

Experimental evaluation of the proposed method, using synthetically generated images, demonstrated that its performance is better than that of a number of other image registration methods for a range of scenes. It was shown that the proposed method is capable of sub-pixel accuracy approaching 1% of a pixel in low-noise situations and 4% of a pixel for moderate noise. The proposed method has been shown to cope well with global illumination variations and is also the fastest, by at least a factor of two, out of the methods tested here. This method would be useful for a range of applications that require fast alignment of shifted images, such as image super-resolution, high-dynamic range reconstruction, panorama stitching.

 ISBN: 978-1-61208-283-7

REFERENCES

[1] A. Goshtasby, *2-D and 3-D image registration for medical, remote sensing, and industrial applications*. Hoboken, NJ: J. Wiley & Sons, 2005.

[2] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[3] A. Rosenfeld and G. J. Vanderbrug, "Coarse-fine template matching," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 7, no. 2, pp. 104–107, 1977.

[4] G. J. Vanderbrug and A. Rosenfeld, "Two-stage template matching," *IEEE Transactions on Computers*, vol. C-26, no. 4, pp. 384–393, 1977.

[5] P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, 1998.

[6] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," *Computer Vision - Eccv 92*, vol. 588, pp. 237–252, 1992.

[7] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK ; New York: Cambridge University Press, 2004.

[8] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artificial Intelligence*, vol. 2, 1981, pp. 674–679.

[9] L. G. Brown, "A survey of image registration techniques," *Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[10] S. Mann, "Compositing multiple pictures of the same scene," in *Proceedings of the 46th Annual IS&T Conference*, vol. 2, 1993.

[11] H. Shekarforoush, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of cross power spectrum," in *Proc. CVPR '96*, Inria,F-06902 Sophia Antipolis,France, 1996, pp. 532–537.

[12] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *IEEE 1975 Conf. Cybernetics and Society*. IEEE, 1975, pp. 163–165.

[13] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration aalgorithm," *Optics Letters*, vol. 33, no. 2, pp. 156–158, January 2008.

[14] E. Vera and S. Torres, "Subpixel Accuracy Analysis of Phase Correlation Registration Methods Applied to Aliased Image," in *16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, August 25-29 2008.

[15] H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, April 1928.

[16] V. N. Dvornychenko, "Bounds on (deterministic) correlation functions with application to registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 206–213, 1983.

[17] K. Turkowski, "Filters for common resampling tasks," in *Graphics gems*, A. S. Glassner, Ed. San Diego, CA, USA: Academic Press Professional, Inc., 1990, pp. 147–165.

[18] T. M. Lehmann, C. Gnner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 1049–1075, 1999.

[19] P. Thévenaz, T. Blu, and M. Unser, "Image interpolation and resampling," in *Handbook of Medical Imaging, Processing and Analysis*. Academic Press, 2000, ch. 25, pp. 393–420.

[20] C. F. Gauss, "Theoria motus corporum coelestium in sectionibus conicis solem ambientium," 1809, Translation: Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections.

[21] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.

[22] H. S. Stone, M. Orchard, E. chien Chang, and S. Martucci, "A fast direct fourier-based algorithm for subpixel registration of images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 2235–2243, 2001.

APPENDIX

Equation 5 describes a resampling filter h_u based on Keys' cubic convolution interpolation kernel. The filter can be used to "shift" a discrete sequence by resampling it on a uniform grid which is offset by u. Because the interpolation kernel has finite support (see Equation 4), the resampling filter has only 4 non-zero coefficients. Assuming $0 \le u < 1$, these four coefficients can be calculated by substituting $m = \{-1, 0, 1, 2\}$ into Equation 5:

$$h_u = [CC(-u-1),\, CC(-u),\, CC(1-u),\, CC(2-u)] \quad (13)$$

To show that Equation 6 is correct, we can substitute the above filter coefficients into the right hand side of this equation and simplify:

$$\begin{aligned}
u =& \sum m h_u(m) \\
=& -CC(-u-1) + CC(1-u) + 2CC(2-u) \\
=& -\left(-\frac{1}{2}|-u-1|^3 + \frac{5}{2}|-u-1|^2 - 4|-u-1| + 2\right) \\
&+ \left(\frac{3}{2}|1-u|^3 - \frac{5}{2}|1-u|^2 + 1\right) \\
&+ 2\left(-\frac{1}{2}|2-u|^3 + \frac{5}{2}|2-u|^2 - 4|2-u| + 2\right) \\
=& \frac{1}{2}(u+1)^3 - \frac{5}{2}(u+1)^2 + 4(u+1) - 2 \\
&+ \frac{3}{2}(1-u)^3 - \frac{5}{2}(1-u)^2 + 1 \\
&- (2-u)^3 + 5(2-u)^2 - 8(2-u) + 4 \\
=& \; u.
\end{aligned}$$

 ISBN: 978-1-61208-283-7