Dornaika, F. & Elder, J.H. (2002). Proc. 7th European Conference on Computer Vision, Copenhagen, in Lecture Notes in Computer Science, vol. 2353, Springer-Verlag, Berlin, 606-620. (2002)

Image Registration for Omnidirectional Sensors

No Author Given

No Institute Given

Abstract

Multi-sensor systems that combine large and small field-of-view (FOV) sensing may be useful in many surveillance and telepresence applications that require both large-field sensing and high resolution at selected locations. This paper addresses the problem of registering high-resolution, small FOV images with low-resolution panoramic images provided by an omnidirectional catadioptric video sensor. Such systems may find application in surveillance and telepresence systems that require a large FOV and high resolution at selected locations. Although image registration has been studied in more conventional applications, the problem of registering omnidirectional and conventional video has not previously been addressed, and this problem presents unique challenges due to (i) the extreme differences in resolution between the sensors (more than a 16:1 linear resolution ratio in our application), and (ii) the resolution inhomogeneity of omnidirectional images. In this paper we show how a coarse registration can be computed from raw images using parametric template matching techniques. Further, we develop and evaluate feature-based and featureless methods for computing the full 2D projective transforms between the two sensors. Our results suggest a superiority of featureless methods for this problem.

Keywords: *vision systems, omnidirectional sensing, foveated sensing, attention, parametric template matching, featureless registration, fusion*

1 Introduction

Over the last ten years there has been increasing interest in the application of omnidirectional sensing to computer vision [3, 10, 12, 18]. Potential applications include surveillance, object tracking, and telepresence [9, 14]. Most existing omnidirectional sensors are catadioptric, i.e. the sensor is composed of a camera and one or more mirrors arranged so that the resulting system has a single viewpoint. Catadioptric sensors allow panoramic images to be captured without any camera motion. However, since a single sensor is typically used for the entire panorama, the resolution may be inadequate for many applications.

There has been considerable work on space-variant (foveated) sensor chips [5, 19]. However, since the number of photoreceptive elements on these sensors is limited, they do not provide a resolution or field of view advantage over traditional chips. Moreover, it is not clear how such sensors could be used to achieve a panoramic FOV over which the fovea can be rapidly deployed. Over the last ten

years, the most common solution to the FOV/resolution tradeoff is to compose mosaics from individual high-resolution images acquired independently by one or more cameras [11, 15, 23]. However, constructing panoramic mosaics is not suitable for real-time applications since it requires capturing a large number of images as well as considerable computation (matching, aligning, merging).

In this paper, we introduce an attentive panoramic sensor conceptually based upon the human foveated visual system, and we describe a framework for automatically combining high-resolution images with low-resolution panoramas provided by an omnidirectional catadioptric sensor. Although image registration has been studied in more conventional applications, the problem of registering omnidirectional and conventional video has not previously been addressed, and this problem presents unique challenges due to (i) the extreme differences in resolution between the sensors (more than 16:1 linear resolution ratio in our application - see Fig. 2 for an example), (ii) the consequent reduction in the number of panoramic pixels within the foveal field-of-view that may be used for registration (less than 0.5% of the raw omnidirectional image), and (iii) the resolution inhomogeneity of omnidirectional images.

The main contributions of the paper are as follows. First, we introduce our attentive panoramic sensor design, which consists of an omnidirectional video sensor and a high-resolution camera mounted on a pan/tilt platform. Second, we show how a coarse registration between the high-resolution images and the low-resolution omnidirectional images can be computed using a parametric template matching technique, using a discrete scale space that can accommodate the inhomogeneity of omnidirectional images. Third, we develop and evaluate two methods, one feature-based and the other featureless, for estimating the 2D projective transform between the high-resolution (foveal) image and the low-resolution omnidirectional images. The featureless approach operates directly on dense image descriptors computed in both images.

The rest of the paper is organized as follows. Section 2 briefly presents our attentive panoramic sensor and the registration problem on which we are focusing. Section 3 describes a coarse registration method based on the use of parametric template matching techniques. Section 4 presents two refinement methods: the first one is based upon point-to-point matching and the robust estimator RANSAC [6]. The second one is based upon a global correlation between image descriptors. Section 5 reports and compares experimental results.

2 The Attentive Panoramic Sensor and Problem Statement

The prototype sensor is shown in Fig. 1(**a**). The panoramic component is a parabolic catadioptric sensor [18]. The parabolic mirror stands roughly two metres from the ground, facing down, and thus images the panoramic field below the ceiling of the laboratory. The foveal component consists of a CCD camera with a 25mm focal length, mounted on a pan/tilt platform. As loaded, the platform travels at an average speed of roughly 60 deg/sec in both pan and tilt

Fig. 1. **(a)** Attentive panoramic sensor. **(b)** Raw foveal image. **(c)** Raw panoramic image. **(d)** Unwarped panoramic image.

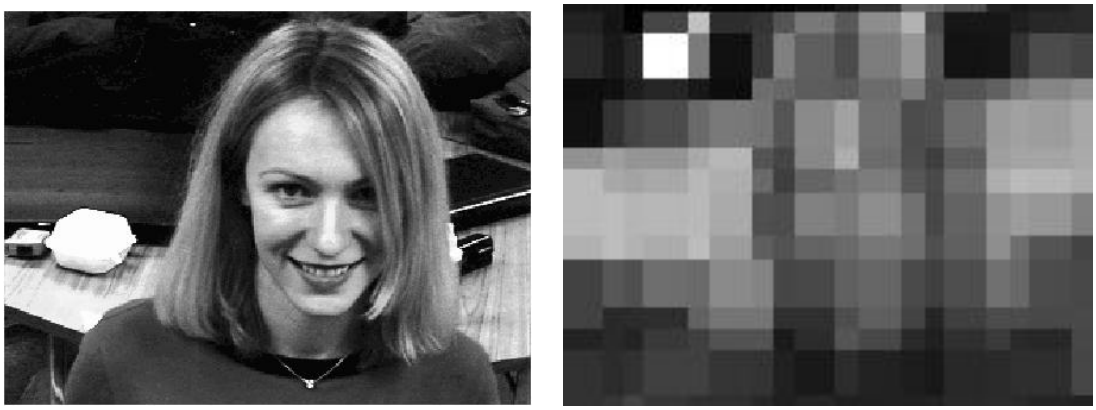


Fig. 2. The foveal image (left) and a (roughly) corresponding region in the panoramic image (right) of Fig. 1.(d)

directions. The vertical axis of rotation coincides with that of the omnidirectional sensor axis. The optical centres of the sensors are separated by 22 cm in the vertical direction. The resolution of the foveal image is 640×480 (Fig. 1(**b**)), the resolution of the raw panoramic image is 640×480 (Fig. 1(**c**)), and that of the unwarped panoramic image is 1024×256 (Fig. 1(**d**)). The field-of-view of the high resolution camera is 14×10 degrees.

The sensor is designed to allow high-resolution video to be selectively sensed at visual events of interest detected in the low-resolution panoramic video stream. These two streams may then be fused and displayed to a remote human observer. Thus each event of interest initiates a three-stage cascade of visual processing:

1. *Stage 1: Detection.* An event of interest is detected in the panoramic stream. In the present system, the event of interest may be selected interactively by a human operator using a point-and-click interface on the fused display, or automatically by a motion detection algorithm.
2. *Stage 2: Saccade.* The pan/tilt platform is used to direct the optical axis of the foveal camera to the (approximate) location of interest. At present a manual calibration procedure is used to construct a lookup table mapping panoramic pixel coordinates to pan/tilt coordinates. During operation, bilinear interpolation on this table generates a pan/tilt command given arbitrary panoramic coordinates.
3. *Stage 3: Fusion.* Foveal pixels are fused with the unwarped panoramic image. This requires (i) an accurate fovea-to-panorama coordinate transform (registration) and a pixel combination rule (fusion). Since the foveal data is of much higher quality than the panoramic data, our fusion rule is essentially to use foveal data where it is available, blending the two smoothly at a roughly circular boundary.

In this paper we address the third stage, and specifically the problem of registration. The problem is made non-trivial by parallax due to the 22cm dis-

placement between the optical centres of the two sensors. To solve this problem we will approximate the mapping between foveal and panoramic images by a 2D projective mapping, i.e. a homography, represented by a 3×3 matrix. This is equivalent to the assumption that within the field-of-view of the fovea, the scene is approximately planar. Solving for the parameters of the projective matrix thus amounts to defining the attitude of the local scene plane. In general, this plane may be different in each gaze direction, and thus for a given static scene one can assume that the mapping between foveal and panoramic coordinates is defined by a 2D (pan/tilt) map of 2D projective matrices.

One possible approach to this problem is to use a manual calibration procedure to estimate these homographies over a lattice of pan/tilt gaze directions, and then to bilinearly interpolate over this table of homographies to estimate an appropriate homography given arbitrary pan/tilt coordinates. At each pan/tilt direction in the lattice, calibration amounts to the selection of at least four pairs of corresponding scene points in panoramic and foveal images, followed by a least-squares estimation of the matrix parameters.

The problem with this approach, as we shall see, is that it works well only for distant or static scenes. For close-range, dynamic scenes, these homographies are functions of time, and so cannot be pre-computed. Thus we require a mapping that is both a function of space (direction in the viewing sphere) and time.

2.1 What Makes Dynamic Registration of Foveal and Omnidirectional Data Challenging?

Several factors makes the automatic registration of foveal and panoramic video streams challenging (Figures 1 and 2):

1. Hundreds of papers have been published on the problems of matching and registration [1]. However, matching and registration are resolution-dependent processes, and these studies generally assume roughly equivalent resolution between sensors. Recent studies that have addressed the problem of resolution differences [4, 8] have considered scale factors up to only a factor of 6. However, in our application, the linear resolution difference between the foveal and panoramic images is as large as 16:1. Moreover, the images registered in previous studies are obtained by the same conventional camera, i.e. the scale differences result solely by optical zooming. Thus the images to be registered are likely to be much more closely related than those obtained by our foveated panoramic sensor.
2. Due to the large resolution difference between fovea and panorama, only roughly 0.5% of the panorama (roughly 50 × 30 pixels) is within the foveal field-of-view. Thus the information available in the panorama for registration is severely limited. The example shown in Fig. 2 suggests that even the human visual system may find this problem difficult with so little information.
3. Unlike conventional images, the resolution of omnidirectional images (provided by catadioptric sensors) varies as a function of viewing direction [2].

For a parabolic mirror, the resolution in the unwarped panorama varies as a function of the elevation, i.e. the vertical coordinate in the panoramic image.

Given the difficulty of the registration problem we have adopted a coarse-to-fine scheme. The registration process is split into two main stages. In the first stage, a coarse registration is computed using a multi-resolution representation of the foveal image and a parametric template matching technique. This provides a rough estimate of the 2D translation and scale factors relating the two images. In the second stage, this estimate is refined into a full 2D projective mapping locally relating the foveal and panoramic images.

3 Coarse Registration

3.1 Multi-resolution Representation

The goal of coarse registration is to estimate a 2D similarity transform between the images consisting of two scale factors (the aspect ratio is not invariant) and a 2D translation. Due to the significant resolution difference, it is difficult to match the foveal image with the panoramic image directly. Instead we employ a discrete Gaussian pyramid representation [13]. The pyramid representation is constructed by smoothing and downsampling the image separably in (x,y) coordinates. While multi-resolution pyramid representations for coarse-to-fine registration have been used before (e.g, [16, 23, 25]), these involve pyramids built from input images of similar resolution, so that that corresponding levels in the pyramids are also matched in resolution. In our work, only one pyramid is built, and the scale factor mapping foveal to panoramic resolution is estimated using parametric techniques.

3.2 Parametric Template Matching over Scale Space

In our system, the scaling factors between foveal and panoramic images are roughly known. The horizontal scale factor is approximately 12:1 for the whole unwarped panorama, and we use this factor in computing the subsampled foveal representation. The vertical scale factor, however, varies from roughly 12:1 to 16:1 within the upper two third part of the panorama, and so a single level of the pyramid will not suffice. We neglect the lower third of the panoramic field of view, since in our system it images primarily the desk on which it stands.

Our approach to this problem is to bracket the expected vertical scale factor with two pyramid levels, one at a scale lower than the expected factor, and the other at a scale higher than the expected factor. Translational mappings between foveal and panoramic images are computed for both scales using conventional template matching techniques, and the optimal similarity transform is estimated parametrically from these. In the following, we first describe our methods for estimating the translational mapping for a fixed level of the foveal pyramid. We then use the parametric technique to estimate the full similarity transform.

3.3 Estimating the Translational Mapping

We employ featureless (correlative) techniques to estimate an approximate 2D translation relating foveal and panoramic images. We have tested several specific techniques including:

1. Minimizing the Sum of Squared Difference (SSD), which is optimal assuming the images differ only by white Gaussian noise
2. Maximizing the correlation, which is near-optimal if the energy of panoramic subimages is approximately invariant
3. Maximizing covariance
4. Maximizing the Pearson correlation (so-called normalized correlation)

We have found the normalized cross-correlation technique to be more reliable than the others, suggesting that the normalization reduces error due to photometric differences between the two sensors.

The normalized cross-correlation between the foveal image $\mathbf{I}_f$ and the panoramic image $\mathbf{I}_p$ at location $\mathbf{p} = (u', v', 1)^T$ is given by:

$$\rho(\mathbf{p}) =< \mathbf{I}_p, \mathbf{I}_f >= \frac{\sum_{(u,v)}(\mathbf{I}_{p(u'+u,v'+v)} - \overline{\mathbf{I}_p})(\mathbf{I}_{f(u,v)} - \overline{\mathbf{I}_f})}{N\,\sigma_p\,\sigma_f} \tag{1}$$

where N is the number of pixels of $\mathbf{I}_f$, $\overline{\mathbf{I}_p}$ is the local average over the overlap region, σ_f and σ_p are the standard deviations associated with $\mathbf{I}_f$ and the panoramic image overlapped by $\mathbf{I}_f$. Thus by maximizing $\rho(\mathbf{p})$ over a part of the panoramic image $\mathbf{I}_p$, we can estimate the 2D translation $\mathbf{p}_r$ between the images:

$$\mathbf{p}_r = \arg\max_{\mathbf{p}} \rho(\mathbf{p}) \tag{2}$$

Since the rotational axis of the foveal camera (the one associated with the pan angle) is coincident with the optical axis of the panoramic camera, and since the optical centre of the foveal camera lies approximately on its rotational axis, the pan/tilt coordinates of the fovea provide information that constrains the probable translation between the images. To quantify these constraints, we determined translational correspondences for a number of scene point pairs at different depths and locations over a regular grid of pan/tilt gaze coordinates. Note that given fixed pan/tilt coordinates, the horizontal mapping is nearly invariant. The vertical mapping varies more, but hard constraints on the sensor system and the room geometry also limits this mapping.

These constraints can be applied by pre-computing means and standard deviations for these distributions over a regular grid of pan/tilt locations. Given pan/tilt coordinates during operation, this table is bilinearly interpolated to estimate the mean and standard deviation of expected translational parameters. We then constrain our panoramic search region to line within the 95% confidence interval determined by these statistics. Typical values for the horizontal and vertical standard deviations are: $\sigma_h = 5$ pixels and $\sigma_v = 10$ pixels.

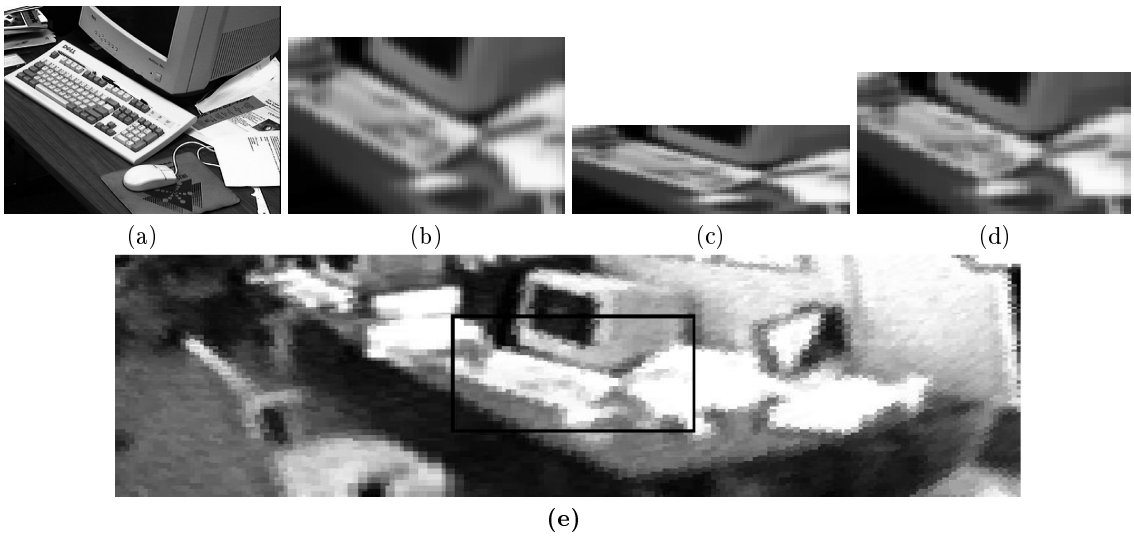


Fig. 3. Coarse registration using parametric template matching based upon two low-resolution representations of the foveal image. **(a)** Original foveal image. **(b)** and **(c)** The reference templates associated with the two vertical scales 1/8 and 1/16, respectively. **(d)** The parametric template associated with the computed vertical scale (0.09 $\approx$ 1/12). **(e)** Panoramic image showing selected location of coarsely-registered foveal parametric template

3.4 Estimating the Full Similarity Transform

Given two computed levels of the foveal pyramid bracketing the true vertical scale factor, we use the parametric template matching method of [24] to estimate the true scale factor relating the foveal and panoramic images given the best translational mapping associated with each reference scale. Fig. 3 shows an example result, using bracketing vertical scale factors of 1:8 and 1:16. The estimated scale factor is 1:12.

4 Fine Registration

The similarity transform computing the coarse registration stage can be used to initialize an estimation of the full local homography relating foveal and panoramic coordinates. Here we develop and compare feature-based and featureless methods for estimating this homography. In both cases, we use the foveal parametric template estimated in the coarse registration phase.

4.1 Feature-based Registration

Feature Matching Our first approach is to accurately localize a large set of simple but reliable features that are very likely to occur in both foveal and

panoramic images, and then to estimate the homography that minimizes a robust measure of correspondence error.

To achieve high localization accuracy and reliability, we select pixels with gradient magnitude, estimated with the Sobel operator, above a specified threshold (we will actually use a multi-threshold method - see below). The 11×11 pixel region centred on each high-gradient pixel is then considered a feature. Given a foveal feature, the expected location of the corresponding panoramic feature is estimated using the similarity transform computed in Stage 1, and this defines the centre of a 13×13 pixel search region in the panoramic image. The panoramic feature with the largest Pearson correlation with the foveal feature is then determined. If the Pearson correlation exceeds a threshold of 0.5, a match is declared.

2D Projective Mapping The local homographic mapping between foveal image coordinates $\mathbf{p} = (u, v, 1)^T$ and panoramic image coordinates $\mathbf{p'} = (u', v', 1)^T$ is given by $\mathbf{p'} \cong \mathbf{H}\,\mathbf{p}$, where $\cong$ represents equality up to a scale factor and $\mathbf{H} \equiv h_{ij}$ is a 3×3 matrix. Thus

$$u' = \frac{h_{11}\,u + h_{12}\,v + h_{13}}{h_{31}\,u + h_{32}\,v + h_{33}} \qquad v' = \frac{h_{21}\,u + h_{22}\,v + h_{23}}{h_{31}\,u + h_{32}\,v + h_{33}} \tag{3}$$

The homography $\mathbf{H}$ can be recovered up to a scale factor using standard linear algebra tools assuming at least four matches. To avoid numerical instability, the image coordinates of the matches in both images are translated and rescaled so that the homogeneous coordinates are close to $(1, 1, 1)^T$.

Robust estimation

Robust estimation techniques provide more reliable estimation in the presence of outliers, in our case, mismatched features [17]. Here we use the RANSAC algorithm [6], which splits the set of putative matches into disjoint outlier and inlier subsets. The estimated homography is that having the largest support among putative matches. Note that if the local scene contains more than one planar surface, the RANSAC paradigm will estimate the dominant plane.

In our approach we further enhance reliability by computing estimates over several gradient thresholds, and then selecting the most consistent. Our developed method combines the use of the RANSAC paradigm with several gradient maps. In other words, the RANSAC paradigm is applied several times allowing the use of more input data (matches). The entire algorithm proceeds as follows.

First, the foveal gradient magnitude map is computed. Using a set of thresholds (10 %, 20 %, 30 %, $\cdots$ of the maximum gradient), a set of thresholded gradient magnitude maps is computed. Next, using the gradient map corresponding to the lowest threshold, matches are computed using the process described above. Then, for each gradient threshold a homography is computed using RANSAC, with only the matched foveal pixels whose gradient exceeds the threshold. Note that the matching process is performed once, at the lowest threshold, and only the linear computation of the homography is performed for each set of matches.

From the resulting set of homographies, the selected homography is that which minimizes the SSD between the warped foveal image and the correspond-

ing panoramic subimage. Table 1 illustrates the application of this heuristic to the foveal image shown in (Fig. 3 (top)) using four different gradient maps. In this example, a gradient threshold of 20% produces the optimal homography.

Threshold	*High gradient pixels*	*Matches*	*Inliers*	*SSD*
10 %	823	264	241	$11.64\ 10^6$
20 %	546	173	169	$11.55\ 10^6$
30 %	392	124	109	$12.15\ 10^6$
40 %	282	98	91	$12.08\ 10^6$

Table 1. The Sum of Squared Differences evaluated with four computed homographies. Each one is computed with a gradient map using the RANSAC technique.

4.2 Featureless Registration

Dense image descriptor Image descriptors are differential greyvalue invariants computed by a combination of image derivatives. These descriptors are invariant to image rotations and translations, and have been used in feature-based approaches for image retrieval and matching [21, 22]. In previous work, each invariant was given by a 7-vector whose components are combination of image derivatives up to third order [21]. In our work, we limit the differential invariant to second order. In other words, the image descriptor at a pixel will be given by the following 4-vector:

$$\mathbf{v} = \begin{bmatrix} L_u\, L_u + L_v\, L_v \\ L_{uu}\, L_u\, L_u + 2\, L_{uv}\, L_u\, L_v + L_{vv}\, L_v\, L_v \\ L_{uu} + L_{vv} \\ L_{uu}\, L_{uu} + 2\, L_{uv}\, L_{uv} + L_{vv}\, L_{vv} \end{bmatrix} \tag{4}$$

where L_i and L_{ij} are the first- and second-order image derivatives, respectively, computed by convolution with Gaussian derivatives. Note that the first component is the square of the gradient norm and the third component is the Laplacian of the local image.

Computing this invariant at every pixel of the foveal parametric template and corresponding panoramic subimage generates eight new images $\mathbf{I}_{fi}, \mathbf{I}_{pi}, i \in [1, .., 4]$, each encoding the spatial distribution of one component of the descriptor $\mathbf{v}$.

2D projective mapping Ideally, if $\mathbf{p}$ and $\mathbf{p}'$ are corresponding pixels of the foveal and panoramic images, the expected values of their image descriptors would be identical. However, to model unknown and possibly varying photometric differences between the sensors, we make the more general assumption that

the image descriptors are related by simple affine transforms:

$$\mathbf{I}_{pi(\mathbf{H}\,\mathbf{p})} = \alpha_i\,\mathbf{I}_{fi(\mathbf{p})} + \beta_i; \quad i \in [1,..,4]$$

Leading to the following objective function:

$$f(\mathbf{H}, \alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3, \alpha_4, \beta_4) = \sum_{i=1}^{4} \sum_{\mathbf{p}} (\mathbf{I}_{pi(\mathbf{H}\,\mathbf{p})} - \alpha_i\,\mathbf{I}_{fi(\mathbf{p})} - \beta_i)^2 \qquad (5)$$

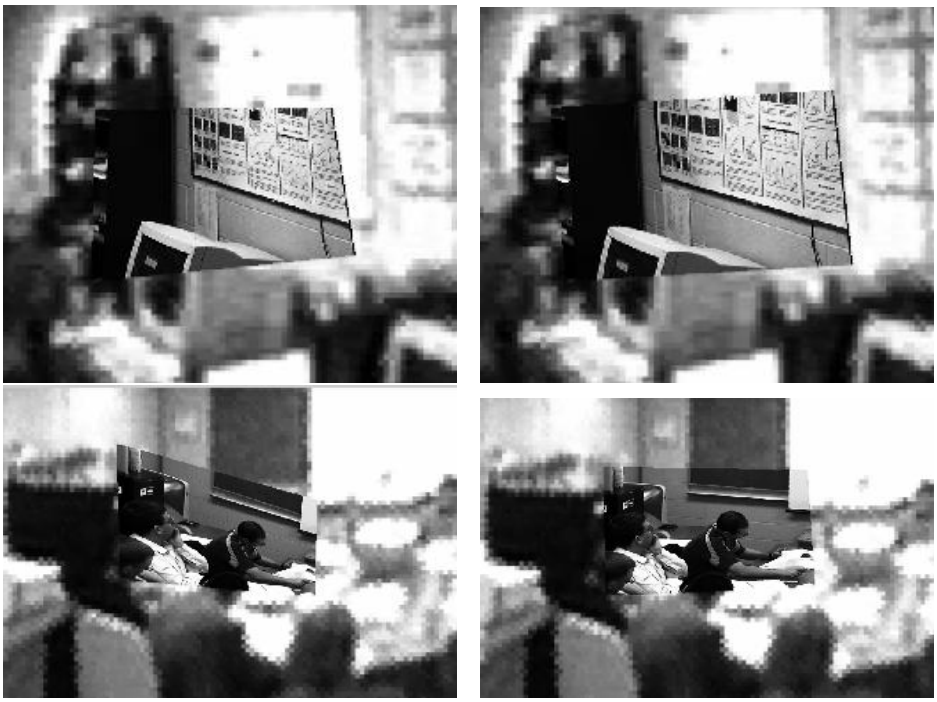
There are sixteen unknowns in this function: the eight entries of the homography matrix (a 2D projective transform has 8 degrees of freedom), and the eight coefficients associated with the four affine mappings. On the other hand, we have $4N$ constraints assuming N foveal pixels. Since the problem involves non-linear constraints, we use the Levenverg-Marquardt technique [7, 20] to estimate the minimum. To increase reliability, we solve in a progressive fashion through increasingly more complex 2D transforms. We start with the similarity transform estimated in the coarse registration stage, and use this to initialize the optimization of a 6-parameter affine mapping. This mapping then initializes the final optimization of the full 8-parameter projective transform.

5 Experimental Results

Fig. 4 shows registration results for three different registration methods: (a) bilinear interpolation of four pre-computed homographies; (b) our feature-matching method Fig. 4 and (c) our featureless method. While both dynamic registration methods improve upon the static calibration, it is clear that the featureless method provides a superior match. The two examples shown in Fig. 5 confirm this finding.

(a) **(b)** **(c)**

Fig. 4. Registration results using three different methods. **(a)** Bilinear interpolation of four pre-computed homographies. **(b)** Feature-matching and robust estimation using RANSAC. **(c)** Featureless method.



(a) (b)

Fig. 5. A comparison between the two registration methods: **(a)** Feature-based registration. **(b)** Featureless registration.

(a) (b) (c)

Fig. 6. Progressive featureless registration **(a)** The coarse registration stage (similarity transform). **(b)** Affine transform **(c)** 2D projective transform. Each stage of computation substantially improves the registration (see the top-left of the fovea).

Fig. 7. Four typical results of featureless registration

Fig. 6 show registration results at three stages of the computation (coarse regisration, affine, projective). Each stage of computation substantially improves the registration. Fig. 5 shows featureless registration for four additional scenes.

The preliminary implementation of the feature-based registration technique takes 0.16 sec on average on an SGI Onyx2, using a panoramic search region of 60×60 pixels. The featureless method takes 0.25-0.5 seconds.

Feature-based vs. featureless registration: quantitative comparison To quantitatively assess the accuracy of the two registration methods, we ran 100 trials of the following simulation over a range of noise levels:

1. A foveal image is subjected to a known, random homography to produce a simulated panoramic image.
2. White Gaussian noise is added to both images, and the homography is estimated using the two registration methods.
3. The transfer error, defined as the average Euclidean distance between the transferred foveal pixels and their ground truth locations is computed.

Fig. 8 displays the results of this simulation. The featureless method is found to outperform the feature-based method at all levels of noise.

6 Conclusion

We have shown that consistent and efficient registration between high-resolution foveal images and low-resolution panoramas provided by an omnidirectional video sensor can

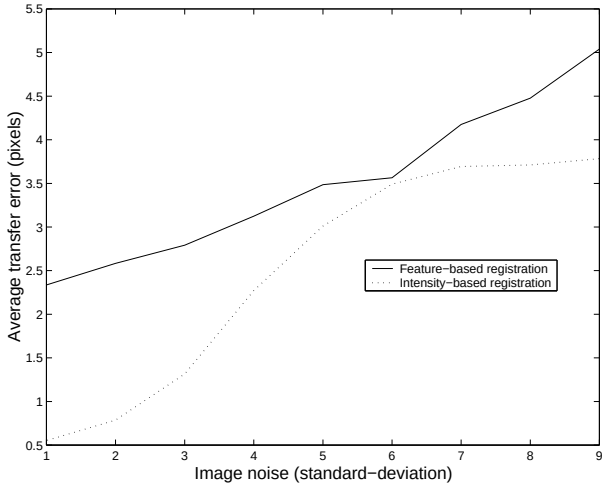


Fig. 8. Average transfer error as a function of image noise for the feature-based (solid) and featureless (dotted) methods.

be achieved. The challenges associated with the extreme resolution differences, the small FOV of the foveal image, and the resolution heterogeneity of the omnidirectional panorama are overcome using a coarse-to-fine scheme. Both feature-based and featureless techniques for registration are developed and quantitatively evaluated, and the featureless approach is found to be superior in all cases. Key elements of our featureless approach are (i) increased robustness by progressively increasing the complexity of the estimated 2D transformation from similarity to affine and projective transformations, (ii) use of the Pearson (normalized) correlation for estimation of the similarity transform, and an affine photometric transform in estimating affine and projective transforms to achieve robustness to photometric differences between sensors, (iii) parametric template matching for rescaling the foveal image to match the heterogeneous panoramic scales, and (iv) transformation of the images to 2D-invariant image descriptor space for robust estimation of affine and projective transforms. These results may be useful for applications in visual surveillance and telepresence demanding both large field-of-view and high resolution at selected points of interest.

References

1. L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, October 1992.
2. T. L. Conroy and J. B. Moore. Resolution invariant surfaces for panoramic vision systems. In *IEEE Conference on Computer Vision*, September 1999.
3. K. Danilidis and C. Geyer. Omnidirectional vision: Theory and algorithms. In *IEEE International Conference on Patter Recognition*, 2000.
4. Y. Dufourneau, C. Schmid, and R. Horaud. Matching images with different resolutions. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.
5. F. Ferrari, J. Nielsen, P. Questa, and G. Sandini. Space variant imaging. *Sensor Review*, 15(2):17–20, 1995.

6. M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communication ACM*, 24(6):381–395, 1981.
7. R. Fletcher. *Practical Methods of Optimization*. Wiley, New York, 1990.
8. B. B. Hansen and B. S. Morse. Multiscale image registrartion using scale trace correlation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1999.
9. I. Haritaoglu, D. Harwood, and L. Davis. Who, when, where, what: A real time system for detecting and tracking people. In *Proceedings of the Third Face and Gesture Recognition Conference*, 1998.
10. R. A. Hicks and R. Bajcsy. Catadioptric sensors that approximate wide-angle perspective projections. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.
11. M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *IEEE International Conference on Computer Vision*, 1995.
12. H. Ishiguro, M. Yamamoto, and S. Tsuji. Omni-directional stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):257–262, 1992.
13. J. Jolion and A. Rosenfeld. *A Pyramid Framework For Early Vision*. Kluwer Academic Publishers, 1994.
14. T. Kanade, R. Collins, A. Lipton, P. Burt, and L. Wixson. Advances in cooperative multi-sensor video surveillance. In *Proceedings of DARPA Image Understanding Workshop*, 1998.
15. R. Kumar, P. Anandan, M. Irani, J. Bergen, and K. Hanna. Representations of scenes from collections of images. In *ICCV Workshop on the Representation of Visual Scenes*, 1995.
16. S. Mann and R. W. Picard. Video orbits of the projective group: A simple approach to featureless estimation of parameters. *IEEE Transactions on Image Processing*, 6(9):1281–1295, 1997.
17. P. Meer, D. Mintz, A. Rosenfeld, and D. W. Kim. Robust regression methods for computer vision: a review. *International Journal of Computer Vision*, 6(1):59–70, 1990.
18. S. Nayar. Catadioptric omnidirectional camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
19. F. Pardo, B. Dierickx, and D. Scheffer. CMOS foveated image sensor: Signal scaling and small geometry effects. *IEEE Transactions on Electron Devices*, 44(10):1731–1737, 1997.
20. W. H. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery. *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press, New York, 1992.
21. C. Schmid and R. Mohr. Local greyvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–534, 1997.
22. C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *IEEE International Conference on Computer Vision*, 1998.
23. R. Szeliski. Image mosaicking for tele-reality applications. In *IEEE Workshop on Applications of Computer Vision*, 1994.
24. K. Tanaka, M. Sano, S. Ohara, and M. Okudaira. A parametric template method and its application to robust matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.
25. Y. Wu, T. Kanade, C.C. Li, and J. Cohn. Image registration using wavelet-based motion model. *International Journal of Computer Vision*, 38(2), 2000.