Registering Conventional Images with Low Resolution Panoramic Images*

Fadi Dornaika

Computer Vision Center
Edifici O, Campus UAB
08193 Bellaterra, Barcelona, Spain
dornaika@cvc.uab.es

Abstract

This paper addresses the problem of registering high-resolution, small field-of-view images with low-resolution panoramic images provided by an panoramic catadioptric video sensor. Such systems may find application in surveillance and telepresence systems that require a large field of view and high resolution at selected locations. Although image registration has been studied in more conventional applications, the problem of registering panoramic 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 panoramic images. The main contributions of this paper are as follows. First, we introduce our foveated panoramic sensor design. Second, we describe an automatic and near real-time registration between the two image streams. This registration is based on minimizing the intensity discrepancy allowing the direct recovery of both the geometric and the photometric transforms. Registration examples using the developed methods are presented.

Keywords: *vision systems, foveated sensing, panoramic sensing, matching, registration, fusion, multi-resolution analysis, non-linear optimization*

1 Introduction

Over the last two decades there has been increasing interest in the application of panoramic sensing to computer vision [1–7]. Potential applications include surveillance, object tracking, and telepresence [8, 9]. Most existing panoramic sensors are catadioptric, i.e. the sensor is composed of a camera and a curved mirror arranged so that the resulting system has a single viewpoint. It has been shown [3] that the projection obtained with a catadioptric sensor with a single viewpoint is equivalent to the projection on a sphere followed by a perspective projection. Catadioptric sensors allow panoramic images to be captured without any camera motion. However, since a single sensor is used for the entire panorama, the resolution of such images may be inadequate for many applications. There has been considerable work on space-variant (foveated) sensor chips [10,

* This work was supported by the MEC project TIN2005-09026.

11]. 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 field of view over which the fovea can be rapidly deployed. A more common solution to the FOV/resolution tradeoff is to compose mosaics from individual overlapping high-resolution images that form a covering of the viewing sphere [12, 13]. These images can be obtained by a single camera that can rotate about its optical centre. Such a system is useful for recording high-resolution ”still life” panoramas, but is of limited use for dynamic scenes, since the instantaneous field of view is typically small. An alternative is to compose the mosaic from images simultaneously recorded by multiple cameras with overlapping fields of view. The primary disadvantage of this approach is the multiplicity of hardware and independent data channels that must be integrated and maintained.

The human visual system has evolved a bipartite solution to the FOV/resolution tradeoff. The field of view of the human eye is roughly 160×175 deg - nearly hemispheric. Central vision is served by roughly five million photoreceptive cones that provide high resolution, chromatic sensation over a five degree field of view, while roughly one hundred million rods provide relatively low-resolution achromatic vision over the remainder of the visual field [14]. The effective resolution is extended by fast gaze-shifting mechanisms and a memory system that allows a form of integration over multiple fixations [15].

In this paper, we introduce our proposed attentive panoramic sensor conceptually based upon the human foveated visual system. More precisely, we propose a framework for automatically combining high-resolution images with low-resolution panoramas provided by a panoramic catadioptric sensor. Although image registration has been studied in more conventional applications, the problem of registering panoramic 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 panoramic image), and (iii) the resolution inhomogeneity of panoramic images. The main contributions of the paper are as follows. First, we introduce our foveated panoramic sensor design, which consists of an panoramic 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 panoramic images can be computed using a parametric template matching technique, using a discrete scale space that can accommodate the inhomogeneity of panoramic images. Third, we develop a fine registration technique for estimating the 2D projective transform between the high-resolution (foveal) image and the low-resolution panoramic images.

The organization of this paper is as follows. Section 2 briefly presents our foveated panoramic sensor and the registration problem on which the paper is focused. Section 3 describes how a coarse registration can be computed using parametric template matching. Section 4 presents a refinement method based upon minimizing intensity discrepancies. Section 5 reports experimental results obtained with our prototype foveated panoramic sensor.

2 Foveated panoramic sensor and problem statement

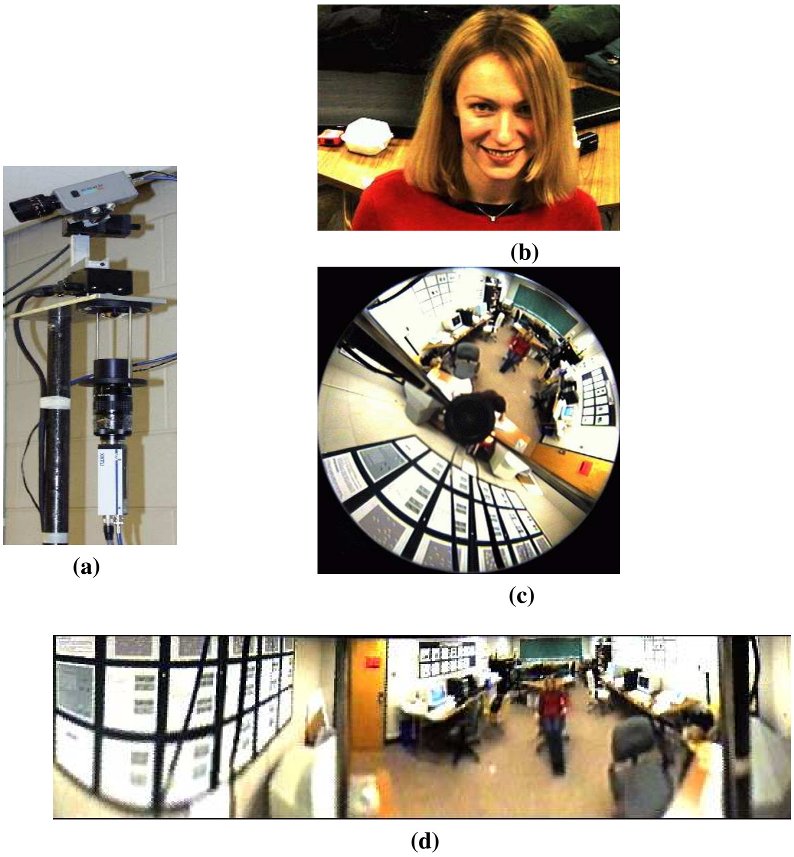


Fig. 1. (a) Foveated panoramic sensor. **(b)** Raw foveal image. **(c)** Raw panoramic image. **(d)** Warped panoramic image.

The prototype sensor is shown in Figure 1(a). The panoramic component is a parabolic catadioptric sensor [6]. 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 colour 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 directions. The vertical axis of rotation coincides with that of the ominidirectional 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 (Figure 1(b)), the resolution of the raw panoramic image is 640×480 (Figure 1(c)), and the resolution of the warped panoramic image is 1024×256 (Figure 1(d)). The field of view of the high resolution camera is 14×10 degrees. The sensor is de-

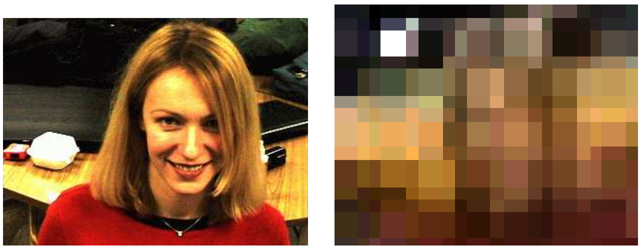


Fig. 2. The foveal image (left) and a (roughly) corresponding region in the panoramic image (right) of Fig. 1.(d).

signed 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.

In this paper we address the problem of registering the high-resolution image with the panoramic one. This problem is made non-trivial by parallax due to the 22cm displacement 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 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. Several factors makes the automatic registration of foveal and panoramic video streams challenging (Figures 1 and 2): (1) In our application, the linear resolution difference between the foveal and panoramic images is as large as 16:1. (2) Only roughly 0.5% of the panorama (roughly 50×30 pixels) is within the foveal field-of-view. (3) Unlike conventional images, the resolution of panoramic images (provided by catadioptric sensors) varies as a function of viewing direction [16].

We will address this challenging registration problem using a coarse-to-fine scheme. The registration process is split into two main stages. In the first stage, a coarse registration is computed using parametric template matching between the panoramic image and a multi-resolution representation of the foveal image. This provides an estimate of the translation and scale factors between the two images. In the second stage, this coarse registration is used to bootstrap a refinement process in which a full 2D projective mapping is computed. This method directly estimates geometric and photometric transforms between the images by minimizing intensity discrepancies.

3 Coarse registration

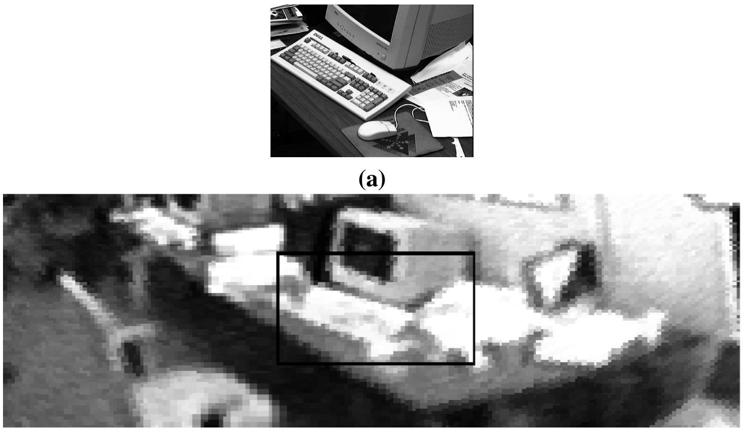
The goal of coarse registration is to find the overlap region in the panoramic image that roughly corresponds to the foveal image. The foveal and panoramic cameras are mounted so that the optical axis of the foveal camera and the *effective* optical axis corresponding to a local patch of the panoramic image are roughly parallel. Thus coarse registration can be achieved by estimating two scale factors[1] and a 2D translation vector, that is, the coarse overlap region is given by a rectangular sub-image. Once this coarse registration is estimated more elaborate methodologies can refine it to a full homography transform (Section 4). Due to the difference in their resolutions, it is difficult to match the foveal image with the panoramic image directly. Instead we employ a discrete Gaussian pyramid representation for the foveal image [17].

Parametric template matching over vertical 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 warped 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 thirds 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 primarily images the desk on which it stands.

Our approach to this problem is to bracket the expected vertical scale with two pyramid levels, one at a scale lower than the expected scale, and the other at a scale higher than the expected scale. Translational mappings between foveal and panoramic images are computed for both scales using conventional template matching techniques, i.e., by maximizing the normalized correlation. Then, the optimal transform (i.e., the vertical scale and the 2D translation) is estimated parametrically from these. Given two computed levels of the foveal pyramid bracketing the true vertical scale factor, we use a parametric template matching method [18] to estimate the true vertical scale factor relating the foveal and panoramic images given the best translational mapping associated with each reference scale. The 2D translation can be computed using the following. Once the vertical scale is estimated, a scaled, low-resolution foveal image is computed from the original high-resolution foveal image, using the estimated vertical scale factor and a scale factor of $1/12$ in the horizontal direction. We then estimate the translational parameters of the coarse registration using normalized cross-correlation of this rescaled foveal image with the panorama.

Figure 3 shows the final coarse registration of the foveal and panoramic images. The translation and scaling transform computed in our coarse registration stage can be used to initialize an estimation of the full local homography relating foveal and panoramic coordinates.

[1] The aspect ratio is not invariant due to the variation in panoramic resolution with elevation.



(a)

(b)

Fig. 3. Coarse registration using parametric template matching based upon two low-resolution representations of the foveal image. **(a)** Original foveal image. **(b)** Panoramic image showing coarse registration of foveal parametric template.

4 Fine registration

In this section we describe how the mapping parameters can be directly estimated from the images without any feature extraction. Our approach involves the direct estimation of mapping parameters by minimization of the discrepancy between the intensity of the two images. Featureless techniques have been applied to the construction of image mosaics in a coarse-to-fine scheme where the 2D transform is iteratively estimated from the coarsest level to the finest level of two pyramids [19]. In this case, the full resolution images as well as the images associated with the two pyramid levels (the low resolution ones) have similar resolution. However, the application of this approach to images of grossly different resolutions has, to our knowledge, not been studied.

We proceed as follows. We denote by $\mathbf{I}_f(\mathbf{p})$ the intensity of the foveal pixel $\mathbf{p} = (u, v, 1)^T$ and by $\mathbf{I}_p(\mathbf{p'})$ the intensity of its match $\mathbf{p'} = (u', v', 1)^T$ in the panoramic image. The image $\mathbf{I}_f$ may be of any resolution including the original (full) resolution.

Foveal and panoramic pixels are assumed to be related by a homography $\mathbf{p'} \cong \mathbf{H}\,\mathbf{p}$, where $\mathbf{H} \equiv h_{ij}$ is a 3×3 matrix such that:

$$u' = \frac{h_{11}\,u + h_{12}\,v + h_{13}}{h_{31}\,u + h_{32}\,v + h_{33}} \tag{1}$$

$$v' = \frac{h_{21}\,u + h_{22}\,v + h_{23}}{h_{31}\,u + h_{32}\,v + h_{33}} \tag{2}$$

Without loss of generality, we set h_{33} to 1 since the homography $\mathbf{H}$ is defined up to a scale factor.

Since these two pixels project from the same scene point, we will assume that their intensities can be related by an affine mapping [20, 21]:

$$\mathbf{I}_p(\mathbf{H}\,\mathbf{p}) = \alpha\,\mathbf{I}_f(\mathbf{p}) + \beta$$

where α is the contrast gain and β is the brightness shift. These parameters cannot necessarily be precomputed, since the sensors may have dynamic gain control.

We thus seek the photometric and geometric parameters of the transformation that minimize

$$f(\mathbf{H}, \alpha, \beta) = \sum_{\mathbf{p}} \nu(\mathbf{p})^2 = \sum_{\mathbf{p}} (\mathbf{I}_p(\mathbf{H}\,\mathbf{p}) - \alpha\,\mathbf{I}_f(\mathbf{p}) - \beta)^2 \tag{3}$$

There are ten unknowns (two photometric parameters, α and β, and the eight entries of the homography matrix), and N non-linear constraints where N is the number of pixels of the foveal image $\mathbf{I}_f$. We use the Levenberg-Marquardt technique [22, 23] to solve this problem. For each foveal pixel, the first derivatives of its contribution to the error function (3) with respect to the ten unknowns have the following form:

$$\begin{aligned}
\frac{\partial \nu}{\partial h_{ij}} &= (\frac{\partial \mathbf{I}_p}{\partial u'}\,\frac{\partial u'}{\partial h_{ij}} + \frac{\partial \mathbf{I}_p}{\partial v'}\,\frac{\partial v'}{\partial h_{ij}}) \quad i, j = 1, 2, 3 \\
\frac{\partial \nu}{\partial \alpha} &= -\mathbf{I}_f(u, v) \\
\frac{\partial \nu}{\partial \beta} &= -1
\end{aligned}$$

where $(\frac{\partial \mathbf{I}_p}{\partial u'}, \frac{\partial \mathbf{I}_p}{\partial v'})^T$ is the spatial gradient vector associated with the panoramic image, and the derivatives, $\frac{\partial u'}{\partial h_{ij}}$ and $\frac{\partial v'}{\partial h_{ij}}$, are easily derived from equations (1) and (2). The Levenberg-Marquardt technique uses these derivatives to iteratively update the transform parameters to minimize the error function.

Due to the complexity of our objective function, it is difficult to obtain a good solution without a good initialization. To increase the reliability of the approach, we estimate the transform in two stages of increasing complexity: first affine (6 parameters) and then projective (8 parameters). For the affine stage, we use as an initial guess the translation and scaling parameters estimated by coarse registration (Section 3). For the projective stage, we use the results of the affine stage as an initial guess. The non-diagonal elements of the initial guess for the affine transform are set to zero. The initial guess for the homography elements h_{31} and h_{32} are set to zero.

5 Experimental results and method comparisons

The featureless registration method described in Section 4 was evaluated over a large number of foveal/panoramic image pairs. Figure 4 shows registration results at three stages of the computation: (**(a)** coarse registration, **(b)** affine, **(c)** projective). Each stage of computation substantially improves the registration.

In our experiments, the 2D projective transform typically provides the best registration. However, we find that for low-contrast foveal images the affine transformation may prove superior. Figure 5 shows such a case. To address such cases, we have developed a post-hoc evaluation technique in which the normalized cross-correlation of both affine and projective transformations of the fovea with the panorama are computed, and the transformation with the largest cross-correlation is selected. In Figure 5, this criterion selects the affine transformation (cross-correlation of 0.77) over the projective (0.57).

Figure 6 demonstrates the benefit of integrating a photometric transform (the parameters α and β) within the optimization process. Objective confirmation of this observation may be obtained by computing the normalized cross-correlations associated with the two transformations. The normalized cross-correlation is greater for the transformation employing both geometric and photometric parameters (0.94) (See Fig. 6(b)) than for the purely geometric transformation (0.90) (See Fig. 6(a)). The average CPU time required for registering the foveal with the panoramic one was about 0.1 seconds including two consecutive non-linear minimizations (affine and projective).

Figure 7 shows registration results for three different registration methods: **(a)** bilinear interpolation of four pre-computed homographies; **(b)** a RANSAC based feature-matching method 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.

6 Discussion

We have shown that consistent and efficient registration between high-resolution foveal images and low-resolution panoramas provided by a panoramic video sensor can be achieved. Although image registration has been studied in more conventional applications, the challenging problem of registering panoramic and conventional video has not previously been addressed. The challenges associated with the extreme resolution differences, the small field-of-view of the foveal image, and the resolution heterogeneity of the panoramic panorama were overcome using a coarse-to-fine scheme. 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. Moreover, the developed registration methods are of general applicability in many fields like remote sensing and video compression. Future work may investigate the enhancement of the featureless registration method by combining an exhaustive and guided search with the gradient descent method.

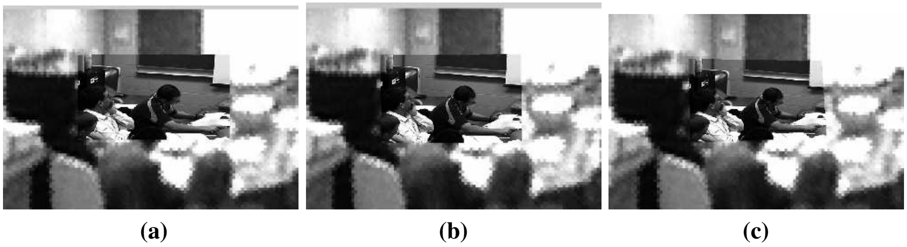


Fig. 4. Progressive featureless registration: **(a)** The coarse registration stage (2 scales and a 2D translation), **(b)** affine transform, and **(c)** 2D projective transform. Each stage of the computation substantially improves the registration (see the top-right of the fovea).

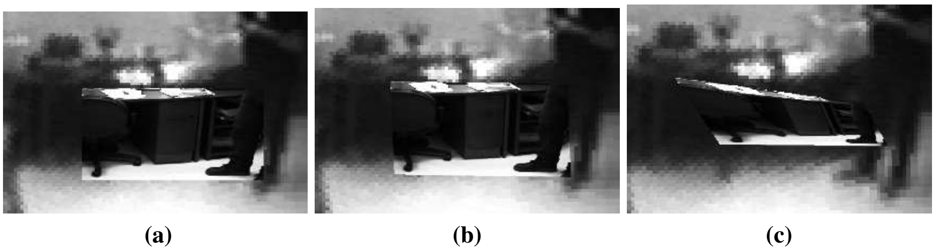


Fig. 5. Progressive featureless registration for a low-contrast foveal image: **(a)** The coarse registration stage (2 scales and a 2D translation), **(b)** affine transform, and **(c)** 2D projective transform.

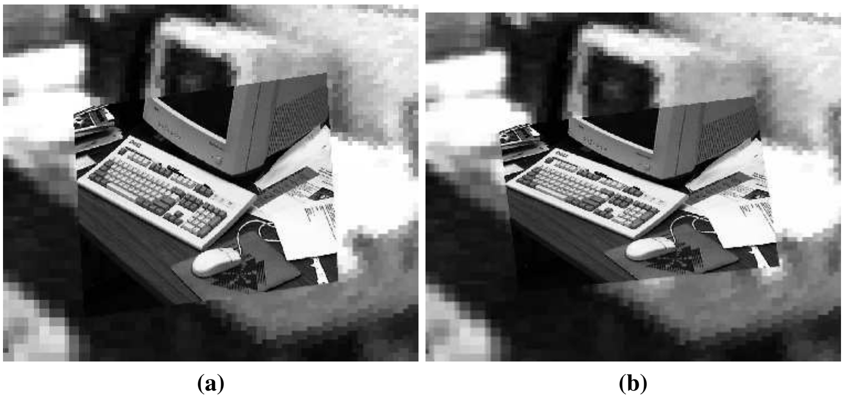


Fig. 6. Featureless registration results. **(a)** Optimization with a purely geometric transform. Note the misregistration of the computer screen and the mouse pad. **(b)** Optimization with a combined geometric and photometric transform.

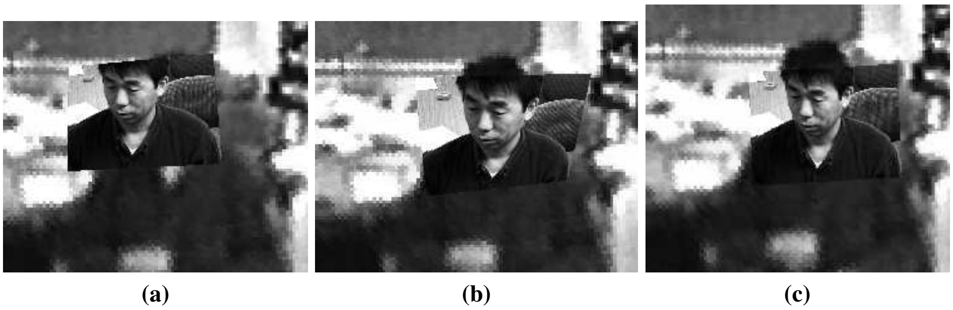


Fig. 7. Registration results using three different methods. **(a)** Bilinear interpolation of four precomputed homographies. **(b)** Feature-matching and robust estimation using RANSAC. **(c)** Featureless method.

References

1. Lin, S.S., Bajcsy, R.: Single-view-point omnidirectional catadioptric cone mirror imager,. IEEE Trans. on Pattern Analysis and Machine Intelligence **28**(5) (2006) 840–845
2. Barreto, J.P.: A unifying geometric representation for central projection systems. Computer Vision and Image Understanding **103**(3) (2006) 208–217
3. Danilidis, K., Geyer, C.: Omnidirectional vision: Theory and algorithms. In: IEEE International Conference on Patter Recognition. (2000)
4. Hicks, R.A., Bajcsy, R.: Catadioptric sensors that approximate wide-angle perspective projections. In: IEEE Conference on Computer Vision and Pattern Recognition. (2000)
5. Ishiguro, H., Yamamoto, M., Tsuji, S.: Omni-directional stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**(2) (1992) 257–262
6. Nayar, S.: Catadioptric omnidirectional camera. In: IEEE Conference on Computer Vision and Pattern Recognition. (1997)
7. Yin, W., Boult, T.E.: Physical panoramic pyramid and noise sensitivity in pyramids. In: IEEE International Conference on Computer Vision and Pattern Recognition. (2000)
8. Haritaoglu, I., Harwood, D., Davis, L.: Who, when, where, what: A real time system for detecting and tracking people. In: Proceedings of the Third Face and Gesture Recognition Conference. (1998)
9. Kanade, T., Collins, R., Lipton, A., Burt, P., Wixson, L.: Advances in cooperative multi-sensor video surveillance. In: Proceedings of DARPA Image Understanding Workshop. (1998)
10. Ferrari, F., Nielsen, J., Questa, P., Sandini, G.: Space variant imaging. Sensor Review **15**(2) (1995) 17–20
11. Pardo, F., Dierickx, B., Scheffer, D.: CMOS foveated image sensor: Signal scaling and small geometry effects. IEEE Transactions on Electron Devices **44**(10) (1997) 1731–1737
12. Kumar, R., Anandan, P., Irani, M., Bergen, J., Hanna, K.: Representations of scenes from collections of images. In: ICCV Workshop on the Representation of Visual Scenes. (1995)
13. Szeliski, R., Shum, H.Y.: Creating full view panoramic image mosaics and texture-mapped models. In: SIGGRAPH'97. (August 1997)
14. Wandell, B.: Foundations of Vision. Sinauer, Sunderland, Massachusetts (1995)
15. Irwin, D.E., Gordon, R.D.: Eye movements, attention and trans-saccadic memory. Visual Cognition **5**(1/2) (1998) 127–155
16. Conroy, T.L., Moore, J.B.: Resolution invariant surfaces for panoramic vision systems. In: IEEE Conference on Computer Vision. (September 1999)
17. Jolion, J., Rosenfeld, A.: A Pyramid Framework For Early Vision. Kluwer Academic Publishers (1994)
18. Tanaka, K., Sano, M., Ohara, S., Okudaira, M.: A parametric template method and its application to robust matching. In: IEEE Conference on Computer Vision and Pattern Recognition. (2000)
19. Mann, S., Picard, R.W.: Video orbits of the projective group: A simple approach to featureless estimation of parameters. IEEE Transactions on Image Processing **6**(9) (1997) 1281–1295
20. Cox, I.J., Roy, S., Hingorani, S.L.: Dynamic histogram warping of images pairs for constant image brightness. In: IEEE International Conference on Image Processing. (1995)
21. Hager, G., Belhumeur, P.: Effecient region tracking with parametric models of geometry and illumination. IEEE Trans. on Pattern Analysis and Machine Intelligence **20**(10) (1998) 1025–1039
22. Fletcher, R.: Practical Methods of Optimization. Wiley, New York (1990)
23. Press, W.H., Teukolsky, S.A., Wetterling, W.T., Flannery, B.P.: Numerical Recipes, The Art of Scientific Computing. Cambridge University Press, New York (1992)