Rotation and Zooming in Image Mosaicing

Fady Badra, Ala Qumsieh, Gregory Dudek

Abstract— **Many methods are available for image mosaicing most of which are not useful because they either (1) require a great deal of overlap between images, or (2) if they work for restricted sub-problem (translation, rotation or zooming) they would not work for the others. While numerous methods exist for accurately calculating translational shifts, no method was found that could handle rotations of angles greater than 15 degrees, or for scaling. In this paper, a new method using Zernike moments is presented to solve for the translational registration, rotational registration(2D and 3D) and zooming all at the same time. This method was tested on different sets of images with translational shifts, rotational shifts and zooming. The method is very fast, efficient and does not require any human interaction, and the results proved to be very accurate.**

I. Introduction

In this paper, we present a new algorithm for an automatic solution for a subset of the general image mosaicing problem. Our approach uses Zernike moments to compute the relative transformations between images. Translational registration, rotational registration(2D and 3D) and zooming can all be recovered simultaneously.

Image mosaicing has numerous applications in medicine, digital terrain mapping and in autonomous navigation of robots. In each of these examples, smaller overlapping images are used to construct an image with a far larger field of view than could be obtained with a single photograph. Potential and existing applications include tasks such as segmentation, object recognition, shape reconstruction, motion tracking, stereo mapping and character recognition. In medical image analysis, relevant research areas include diagnostic medical imaging, such as tumour detection and disease localization and biomedical research including classification of microscopic images of blood cells, cervical smears, and chromosomes. Registration methods can also be applied to civilian and military applications, agriculture, geology, oceanography, oil and mineral exploration, and target location and identification.

Zernike moments belong to a set of orthogonal polynomials, which allow independent moment invariance to be constructed easily to an arbitrarily high order. Such moments have several advantages in terms of noise sensitivity, information redundancy and image representation ability. Thus, Zernike moments offer a complete solution to the recovery rotational and scaling parameters without the need for extensive correlation and search algorithms.

F. Badra is a graduate student at the Virtual Reality (VR) Group, McGill University, Montreal, Canada. e-mail:badraf@ee.mcgill.ca

A. Qumsieh is a graduate student at the Center for Intelligent Machines (CIM), McGill University, Montreal, Canada. e-mail:qumsieh@cim.mcgill.ca

G. Dudek is an associate professor at the Center for Intelligent Machines, McGill University, Montreal, Canada. e-mail:dudek@cim.mcgill.ca

The next section presents a detailed review of the related work. In section III, we discuss some previous work and background on the problem of mosaicing. Our approach is discussed in section IV, while section V demonstrates some of the results. As a conclusion, we investigate the robustness and significance of our method along with a discussion of potential future applications.

II. Background

The main problem is to find the overlapping region between two neighbouring images and to paste them together and can be summarized as follows. Given a sequence of spatially overlapping images, find the transformation between adjacent images, align them, and use the results to construct a larger compound view (See Figure 1). The process of finding the relative transformation between any two images is called *image registration*, while the process of generating panoramic 3D views is called *image mosaicing* or *stitching*.

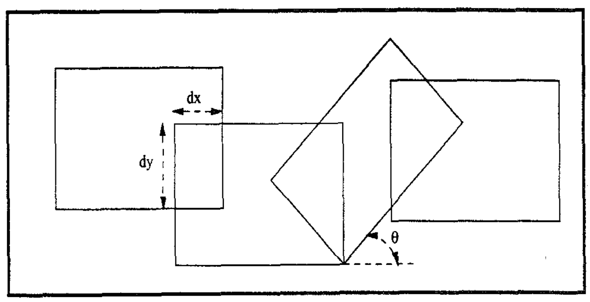

Fig. 1. General mosaicing problem

There is no direct method available for automatically preparing image mosaics. The simplest approach used in practice to prepare a mosaic is to assemble the photographs manually. The key problem in automating the process lies in developing a good algorithm to optimally determine the seam between images of neighbouring regions. In essence, this is an instance of the correspondence problem. The problem of spatial assembling of these images is solved manually by selecting a few points of interest in the overlapping region, setting the correspondence manually and then obtaining the necessary transformation.

Efficient methods have been developed to automatically solve for the translational shift. In the case when there is a large overlap between the images, Szeliski[7] has proposed a method for image registration by directly minimizing the discrepancy in intensities between pairs of images. An optimal solution is found using the Levenberg-Marquardt nonlinear minimization algorithm. This method has the advantage of not requiring any feature points and being sta-

0-8186-8606-5/98 $10.00 © 1998 IEEE

tistically optimal in the vicinity of the true solution, but it needs *a priori* approximation of the relative transformation between the images. In practice, the Levenberg-Marquardt algorithm is computationally expensive, and does not necessarily converge to the desired minimum.

On the other hand, if the overlap between the images is small, a good method was devised where hierarchical matching is used to avoid local minima[10]. Kuglin and Hines[3] proposed another method called *phase correlation*, which gives good results when the camera motion is very large. This method relies on the fact that a translation in the spatial domain corresponds to a phase shift in the frequency domain. Although the results are quite accurate, phase correlation requires a memory capacity that grows with the log of the image area, which makes it slow when handling larger images.

In the case of rotational shift, few methods proved to be efficient. Among the best methods we have the work of Dani and Chaudhuri[2]. Their general method consists of three stages. In the first stage, features in the image are computed. In the second stage, feature points in the reference image, often referred to as control points, are corresponded with feature points in the data image. In the last stage, a spatial mapping is determined using these matched feature points. Reassembling of one image onto the other is performed by applying the spatial mapping and interpolation. This method works well for up to 15 degrees of rotation using angles between edge points.

De Castro and Morandi[1] used the fact that rotation is invariant with the Fourier transform; rotating an image rotates the Fourier transform of that image by the same angle, ϕ. If we know the angle, then we can rotate the cross-power spectrum and determine the translation by phase correlation. However, since ϕ is not known, the phase of the cross-power spectrum is computed as a function of the rotation angle estimate. By first determining the angle ϕ which makes the inverse Fourier transform of the phase of the cross-power spectrum the closest approximation to an impulse, the translation is determined simply as the location of the pulse. Applying the method to a zero-padded image showed that it is very costly because of the difficulty in testing for each angle ϕ.

Viola[4] used mutual information method for images with 30 degrees rotation, but it also required a large amount of overlap, larger than 50%. Zoghlami and Faugeras[5] used corners to build a 2D mosaics from a set of images. In addition to requiring an overlap of at least 50%, the method could successfully recover rotations only around the optical axis of the camera.

Peleg and Herman[6] used manifold projections to generate a panoramic view. This method is well suited to images with small degrees of rotations and does not address the zooming problem. While Rousso, Peleg and Finci[11] devised a method to create image mosaics using generalized strips. Basically, strips of the images, that are perpendicular to the optical flow pattern, are extracted. These strips are then warped to ensure the continuity of the image, and finally pasted successively. Results using this technique are quite satisfactory, but require the precomputing of optical flow fields, which are usually noisy and not very reliable. In addition, flow computation assumes dense temporal sampling. Small rotations are accounted for, but large rotations can not be handled.

Thus, existing mosaicing methods have strong limitations on imaging conditions and distortions are very common. No method was given that could simultaneously solve for lateral translation, rotation about any point, *and* zooming.

In attempting to propose a good solution, we make the following assumptions:

- There is no warping of images (no elastic deformation).
- The subimages do not exhibit a high level of repetitive similarity, for example, a sinusoidal or a checker board variation in intensity.
- Intensity in the image does not change arbitrarily, but there may be a change in overall contrast due to changes in illumination or camera parameters.
- There is at least 10% overlap between adjacent images.

The first of these assumptions indicates that our proposed algorithm would not work for higher order alignment models, such as affine or perspective. That is a true limitation of our current implementation, as well as most other image mosaicing techniques. But, we believe there is room to extend our approach further to deal with these issues. Yet, as the algorithm stands, we see two cases where our technique will prevail over other implementations:

- A camera mounted on a tripod. In this case, the camera is not allowed to change its position in space. It is only able to translate in any direction, and rotate about its vertical axis. Therefore, no warping could occur.
- A handheld camcorder in which the motion is slow and uniform. In this case, slight warping might occur, but can be safely neglected due to the large overlap between consecutive images.

In the case where there is no overlap of the scenes in two such images, the registration cannot be accomplished without *a priori* knowledge of the pictured scene. Heuristic approaches such as analytic continuation of curvilinear elements in spatially adjacent photographs may be used under such circumstances, but this is beyond the scope of this paper.

This article presents a new method to compute the registration function between two images requiring only very little overlap ($10-20\%$) and arbitrary rotation. Our technique will also take care of magnification and zooming. It is very fast, completely automatic without any human interaction and can handle any number of images.

III. Theory

A. Fourier Transforms

Kuglin and Hines[3] proposed a method to align two images which are shifted relative to one another. This method is called phase-correlation, and relies on the translation property of the Fourier Transform; basically, a lateral shift in spatial domain translates into a phase shift in the Fourier domain. The method states that given two images f_1 and f_2 which differ only by a shift (dx, dy), their respective Fourier Transforms can be found

$$F_1 = |F_1|e^{-i\phi_1} \tag{1}$$

$$F_2 = |F_2|e^{-i\phi_2} \tag{2}$$

Multiplying F_1 by $F_2{}^*$, and dividing by the magnitudes, results in a function that depends only on the difference of the phases of the two input images

$$\begin{aligned} \Delta &= \frac{|F_1||F_2|e^{-i(\phi_1-\phi_2)}}{|F_1||F_2|} \\ &= e^{-i(\phi_1-\phi_2)} \end{aligned} \tag{3}$$

Now, taking the inverse Discrete Fourier Transform of Δ results in a delta function whose peek corresponds to the relative spatial shift between the two images.

Although this method does not work very well in the case of different scaling, and large rotations, it still gives good results for small amounts of rotation. In the ideal case where the relative transformation between the images is a pure translation, phase-correlation proved to be extremely accurate to within a single pixel.

B. Zernike Moments

Zernike Polynomials constitute a complete set of complex polynomials that are orthogonal over the interior of the unit circle, $x^2 + y^2 = 1$. Since these polynomials are complex, and denoting them by $V_{nm}(x, y)$, they can be expressed by a radial and an angular components

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho)e^{im\theta} \tag{4}$$

where
n is a non-negative integer,
m is an integer obeying the constraints $n - |m|$ is even and $|m| \leq n$,
ρ is the magnitude of the vector from the origin to the point $(x, y) = x^2 + y^2$,
θ is the angle between the vector and the x-axis, in a counterclockwise direction.
The Radial component $R_{nm}(\rho)$ is defined as

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!}\rho^{n-2s} \tag{5}$$

It should be noted that the radial polynomials are symmetric with respect to m, i.e. $R_{nm}(\rho) = R_{n,-m}(\rho)$. The orthogonality of the Zernike Polynomials is expressed by the following relationship

$$\int\int_{x^2+y^2\leq1} V^*_{nm}(x,y)V_{pq}(x,y)\ dx\ dy = \frac{\pi}{n+1}\delta_{np}\delta_{mq} \tag{6}$$

where

$$\delta_{ab} = \begin{cases} 1 & a = b \\ 0 & otherwise \end{cases} \tag{7}$$

The projection of the image function $f(x, y)$ onto this orthogonal set of polynomials define the Zernike moments. More specifically, the Zernike moment of order n, with repetition m, has the form

$$A_{nm} = \frac{n+1}{\pi}\int\int_{x^2+y^2\leq1} f(x,y)V^*_{nm}(\rho,\theta)\ dx\ dy \tag{8}$$

In the case of a digital image, which is composed of a discrete number of pixels, the double integral simply reduces to a double sum

$$A_{nm} = \tfrac{n+1}{\pi}\textstyle\sum_x\sum_y f[x,y]V^*_{nm}(\rho,\theta) \quad , \quad x^2+y^2 \leq 1 \tag{9}$$

Note that the equation of the Zernike moment holds only within the unit disk. In order to correctly compute the Zernike moments of an image, the image has to be mapped first onto the unit circle. Only pixels lying within the unit circle are used.

Once the Zernike moments are computed, the orthogonality of the Zernike basis allows the use of these moments to reconstruct the image. In order to perfectly recover the input image, an infinite number of moments has to computed. Since this is not possible, only a discrete replica $\hat{f}(x, y)$ can be recovered by computing the moments up to a given moment n_{max}. $\hat{f}(x, y)$ is computed as follows

$$\hat{f}(x,y) = \sum_{n=0}^{n_{max}}\sum_m A_{nm}V_{nm}(\rho,\theta) \tag{10}$$

Now consider an image $f(r, \theta)$, and a rotated replica $f_r(r, \theta)$ rotated by an angle α. Those two images are related as such

$$f_r(r,\theta) = f(r,\theta-\alpha) \tag{11}$$

By the virtue of their definition, Zernike moments of both of these images are coupled by a simple relationship. In polar coordinates, where $x = \rho\cos\theta$ and $y = \rho\sin\theta$, we can express the Zernike moments of the original image as

$$\begin{aligned} A_{nm} &= \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 f(\rho,\theta)V^*_{nm}(\rho,\theta)\rho\ d\rho\ d\theta \\ &= \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 f(\rho,\theta)R_{nm}(\rho)e^{-im\theta}\rho\ d\rho\ d\theta \end{aligned} \tag{12}$$

The equations for the rotated image are

$$A^r_{nm} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 f(\rho,\theta-\alpha)R_{nm}(\rho)e^{-im\theta}\rho\ d\rho\ d\theta \tag{13}$$

Letting $\beta = \theta - \alpha$, we can relate both sets of Zernike moments as follows

$$\begin{aligned} A^r_{nm} &= \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 f(\rho,\beta)R_{nm}(\rho)e^{-im(\beta+\alpha)}\rho\, d\rho\, d\beta \\ &= \left[\frac{n+1}{2\pi}\int_0^{2\pi}\int_0^1 f(\rho,\beta)R_{nm}(\rho)\cdot e^{-im\beta}\rho\, d\rho\, d\beta\right] \\ &\quad \times e^{-im\alpha} \\ A^r_{nm} &= A_{nm}e^{-im\alpha} \end{aligned} \quad (14)$$

Equation 14 shows that, under simple rotation, Zernike moments merely acquire a phase shift. Therefore, if the complex values of A_{nm} and A^r_{nm} could be computed, then the relative angle of rotation α between the two images can be easily recovered. Equation 14 also has an additional advantage. By taking the ratio of the magnitudes of A_{nm} and A^r_{nm}, the relative scaling, s, of the two images can be approximated to within a reasonable degree of accuracy

$$s = \frac{|A_{nm}|}{|A^r_{nm}|} \quad (15)$$

IV. Algorithm Details

Our approach to image registration is based on selecting a preliminary correspondence, and then refining it using the Zernike moment computation. If the initial correspondence proves incorrect, an alternative correspondence is attempted. The process is summarized in the following pseudo-code, followed by a detailed explanation.

1. Select potential correspondences using interest operator, define points of points p_1 and p'_1.
2. Calculate initial estimate $t_0 = (p_{1x} - p'_{1x}, p_{1y} - p'_{1y})$
3. Compute Zernike moments about p_1 and p'_1.
4. Estimate R_n and S_n.
5. Align Images.
6. Compute t_{n+1} using FFT.
7. if $|t_{n+1} - t_n| < \delta \rightarrow$ END, else go back and recompute moments (goto step 3)

As noted above, the iterative solution for both translation and rotation is based on an initial estimate $t_0 = (p_1 - p'_1)$ to define a common center of rotation; that is, a pair of points that are projections of the same point in the scene. The alignment of the projections of such points provides a preliminary estimate of translation. It should be noted, however, that due to the recursive nature of the algorithm, the chosen pairof points *need not* correspond to the *exact* same physical location. Zernike mo ments are robust enough to give accurate results even when there is a difference between the corresponding points. Prior work on feature-based correspondence has used a set of corner points under this type of assumption to solve to entire mosaic problem [5]. Corner points, however, are based on specific assumptions regarding the scale and structure of distinctive image content. We have evaluated the use of an alternative "interest operator" to define a small number of points which are candidates for preliminary correspondence.

We propose the use of an attention operator inspired by observations of human visual saccades [14]. Visual saccades are drawn to regions of unusual image content, and in particular regions with anomalous edge density. Similarly, a robust attention operator can be defined computationally which defines interest points as those regions of the image whose local edge-element density is maximal over the image [15]. While a detailed consideration of this operator is outside the scope of this paper, it appears to be robust and effective over a wide range of scene types.

A disk of a predefined radius is mapped around each of the points, and the Zernike moments within each disk are computed. Initial estimates, R_0 and S_0, of the relative rotation and scaling of the two images are then calculated. This allows one of the images to be brought into correspondence with the other by applying the suitable scaling and rotation. Given this correction for scale and rotation, we can refine the translation estimate t_0, to produce a more precise estimate t_1. This is computed from phase correlation of the underlying images, a technique with several desirable robustness properties [13]. If the refined translation estimate differs substantially from the previous one, the scaling and rotation is re-evaluated. This process repeats until the two images align perfectly, or until the alignment error is below a threshold δ.

The program iterates until a suitable transformation is recovered. The termination criterion for the iteration process is satisfied when the difference between successive estimates of the lateral shift is below a certain threshold. Accurate results can be achieved if the threshold is lowered and the program is left to iterate further. On the other hand, resolution could be traded off for speed by simply increasing the threshold, and performing fewer iterations.

V. Results and Discussion

This section discusses the validity and robustness of our algorithm. It also presents some of the panoramic mosaics that were automatically generated using our approach. As can be seen later on, our algorithm can easily handle images with little overlap, large rotations(2D/3D) about any axis, and scale differences.

A. Robustness

In order to verify the robustness of our algorithm, a set of tests was devised. A set of images with different scales and rotations were generated. These values were chosen so as to test the limitations of using Zernike moments for rotation and scale recovery. Figures 2 and 3 show the results.

In figure 2, close examination shows that Zernike moments gave almost accurate results for *any* angle of rotation. Only shown are values corresponding to positive angles. Since negative angles can be thought of as reversing the order of the images, Zernike moments can recover any angle in the range of $-\pi$ to π.

In figure 3, we can see that the recovered values were very accurate for a scale factor of $\pm70\%$. Beyond that limit, Zernike moments gave erroneous results that varied

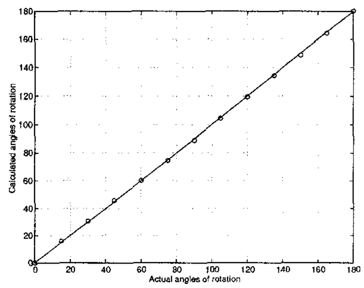

Fig. 2. Plot of the actual angles of rotation versus the calculated ones. All angles are in degrees.

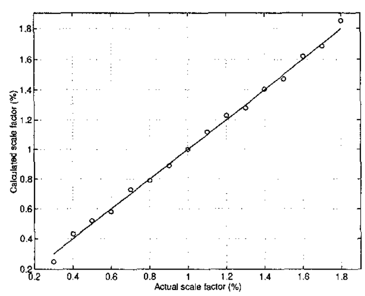

Fig. 3. Plot of the actual scale factors versus the calculated ones.

with the contents of the input images.

B. Translation and Rotation

Various attempts by others to solve the rotation problem proved successful for only "small rotations" [11], while others assumed a rotation only about the optical axis of the camera[5]. Here, we show that our algorithm can successfully recover *any* angle of rotation about *any* axis. The trick is that once a corresponding pair of points has been identified in the two images, this point is taken as the center of rotation. Figures 4 and 5 show some results.

The relative rotation between the images in figures 4 and 5 range between 35 to 50 degrees. In figure 4, the relative rotation between the last two images is exactly 42 degrees. The algorithm computed this angle and found it to be 41.89 degrees. While in figure 5, the relative rotation between the two images is unknown. The algorithm computed this rotation to be 47.03 degrees. As apparent from the final images, our algorithm is extremely accurate for any angle of rotation, and the results are almost perfect.

C. Zooming

Other known mosaicing methods assumed no difference in scale among images. Among the most elaborate solutions, [11] provides only "preliminary" results. In this section, we present results for the zooming problem using our method (See equation 15). We show that our algorithm can efficiently and accurately recover zoom scales between images of up to 150%. Moreover, the algorithm can detect zoom scales even if there is a relative rotation between the images. In fact, both the rotation angle and the zoom scale are computed *at the same time.*

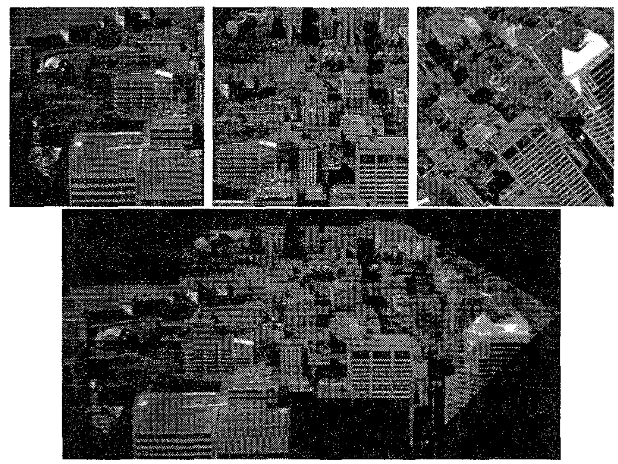

Fig. 4. Shown above are the three original images. The translation and phase difference between them is apparent. The resulting mosaic is shown on the bottom.

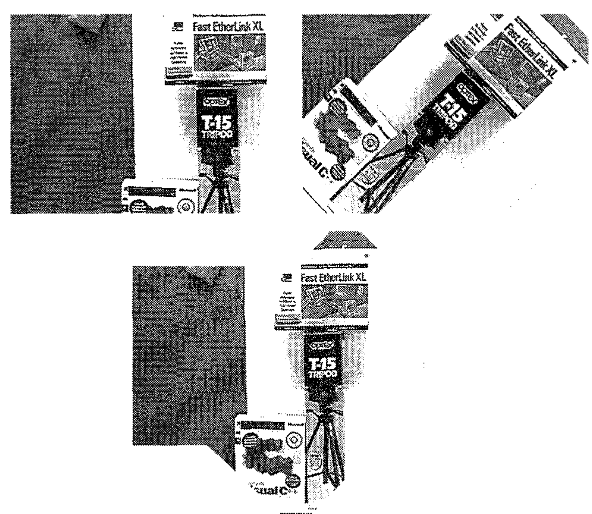

Fig. 5. Another mosaic involving both rotation and translation.

During dilation, the field of view becomes smaller; while, during zooming out, the field of view becomes larger. This creates a difference in resolution between images of different scales. In order to minimize the difference in resolution between images, all images are scaled down to the size of the smallest one, before being stitched. Linear and fractal interpolations can also be used to reduce this resolution discrepancy.

In figure 6, The two images are of the same size. There is a difference factor of 2 between the relative scales. Only parts of the images bounded by the boxes were used to generate the final mosaic on the left. The reason for cropping the images is that, due to the scale difference, one of the images is totally contained in the second one. As can be seen, the algorithm correctly detected the scale difference.

VI. Conclusions and Future Work

In this paper, we have introduced another approach to the image mosaicing problem. The presented algorithm could quickly and efficiently solve for the relative translation, rotation and scaling between the input images. The quality of the final images is extremely good, and discrepancies at image boundaries are nonexistent. The algorithm is extremely fast, and could generate mosaics in real time. It could be used with a robot, where images taken by the robot are quickly stitched together as they are obtained. One limitation of our current implementation of the algorithm is its lack of support for higher order transformations.

The algorithms utilized Zernike moments to estimate the rotation and scaling differences among the images. It has been shown in studies comparing several types of moments (regular, complex, Legendre, Zernike, pseudo-Zernike) that Zernike moments are superior to other types of moments in terms of image representation capability, information redundancy and noise sensitivity [12].

Future application of image mosaics could be in dynamic virtual reality(sometimes called telepresence), which composites video from multiple sources in real time to create the illusion of being in a dynamic(and perhaps reactive) 3D environment. An example application might be to view a 3D version of a concert or a sporting event with control over the camera shots. Another application could be virtual classroom. Building such dynamic 3D models at frame rates is beyond the processing power of today's high performance workstations, but it could be achieved using a collection of these machines or a special purpose stereo hardware.

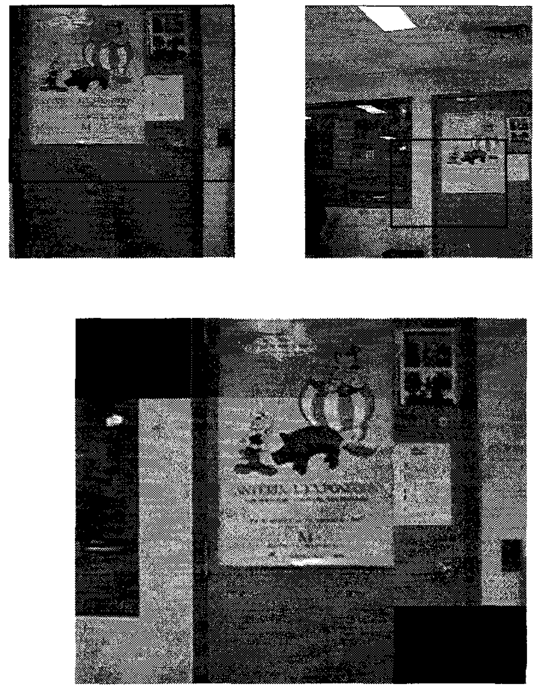

Fig. 6. A panoramic image showing scaling. The two images on the left are the original images. Only the subimages shown inside the box were used to generate the final image on the right.

References

[1] E. D. Castro and C. Morandi, *Registration of translated and rotated images using finite Fourier transform*, IEEE Trans on Pattern Analysis and Machine Intelligence, PAMI-9 ,700–702, 1997.

[2] P. Dani and S. Chadhuri, *Automated assembling of images: Image montage preparation*, Pattern Recog., 28(3):431–445, 1995.

[3] C. D. Kuglin and D.C. Hines, *The phase correlation image alignment method*, In Conference on Cybernetics and Society, pp. 163–165, New York, 1975.

[4] P. Viola, *Alignment by Maximization of mutual information*, TR 1548, MIT, MIT Department of Electrical Engineering and Computer Science, Cambridge, MA, 1995.

[5] I. Zoghlami, O.Faugeras and R. Deriche, *Using geometric corners to build a 2D mosaic from a set of images*, Computer Vision and Pattern Recog. pp 420–425, Puerto Rico, June 1997.

[6] S. Peleg and J. Herman, *Panoramic Mosaics by Manifold Projection*, Computer Vision and Pattern Recog. pp 338–343, Puerto Rico, June 1997.

[7] R. Szeliski, *Image mosaicing for tele-reality applications*, Technical Report CRL 94/2, DEC-CRL, May 1994.

[8] L. G. Brown, *A survey of image registration techniques*, ACM computing Surveys, Vol. 24, No. 4, pp. 325–376, December 1992.

[9] W.H. Press et al., *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edition, Cambridge Univ. Press, Cambridge, England, 1992.

[10] A. Witkin, D.Terzopoulos, and M.Kass, *Signal Matching through Scale Space*, International Journal of Computer Vision, Vol. 1, pp. 133–144, 1987.

[11] B. Rousso, S. Peleg, and I. Finci, *Mosaicing with Generalized Strips*, DARPA Image Understanding Workshop, May 1997.

[12] C. H. Teh and R. T. Chin, *On Image Analysis by Methods of Moments*, IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-10, No. 4, pp. 496–513, July 1988.

[13] D. J. Fleet and A. D. Jepson, *Computation of Component Velocity from Local Phase Information*, International Journal of Computer Vision, IJCV-5, No. 1, pp.77–104, August 1990.

[14] D. Noton and L. Stark, *Eye Movements and Visual Perception*, Scientific American, Vol. 224, No. 6, pp. 33-43, June 1971.

[15] R. Sim and G. Dudek *Navigation by the Stars: A Method for Vision-based Robot Localization*, CIM Technical Report, CIM TR-7634, September, 1997.