Superresolution technique for planar objects based on an isoplane transformation

Yeol-Min Seong
HyunWook Park, MEMBER SPIE
Korea Advanced Institute of Science and Technology (KAIST)
Department of Electrical Engineering
373-1 Guseong-dong, Yuseong-gu
Daejeon 305-701, Korea
E-mail: hwpark@ee.kaist.ac.kr

Abstract. This paper presents an automatic superresolution method using multiviewpoint images. Viewpoints of objects in consecutive image frames often change in real video sequences. Hence, in order to adapt the conventional superresolution methods to multiviewpoint images, a geometric transformation of multiviewpoint images to a reference image plane should be performed. In this process, an accurate isoplane transformation is required. We propose a robust random sample consensus (RANSAC) criterion and a weighted homography estimation, which are important for accurate geometric transformation. Experiments were performed with several video sequences that simulate real surveillance systems. Multiviewpoint image sets were also used to verify the accuracy and stability of the proposed methods. The experimental results show that low-resolution images that are difficult to discern become recognizable high-resolution images using the proposed superresolution method.
 [DOI: 10.1117/1.2931461]

Subject terms: isoplane transformation; multiviewpoint images; RANSAC; superresolution; weighted homography.

Paper 070684R received Aug. 11, 2007; revised manuscript received Jan. 15, 2008; accepted for publication Mar. 5, 2008; published online May 22, 2008.

1 Introduction

Surveillance systems have become more important in recent years because they play a critical role for security when and where an operator cannot monitor. However, surveillance systems often do not provide enough information in terms of image quality and resolution. Hence, resolution enhancement is needed to improve the image resolution. Many interpolation methods have been developed for resolution enhancement, but they have inherent restrictions on the amount of information available. In order to utilize more information, superresolution methods reconstruct a high-resolution image using multiple low-resolution images.[1,2]

Accurate subpixel registration between low-resolution images is important for the superresolution methods. However, subpixel registration errors may exist in the processing of real video sequences. Lee and Kang[3] proposed a superresolution method considering inaccurate subpixel registration: They chose a regularized iterative reconstruction algorithm to solve the ill-posed problem. Many superresolution methods, including Lee and Kang's method,[3] assume that there are only translational shifts among the low-resolution images. However, rotation and scaling between image frames frequently occur in real image sequences. Therefore, we combine multiviewpoint image processing with superresolution reconstruction in order to consider rotation and scaling, where projective transformation (2-D homography) is adopted for the multiviewpoint image processing.[4,5]

This paper presents a robust random sample consensus (RANSAC) and weighted homography estimation, which reduce the distortions of the isoplane transformation. The proposed method considers the confidence level as well as the geometric information of the matching features.

The remainder of this paper is organized as follows. The proposed superresolution technique with accurate isoplane transformation is described in Sec. 2. In Sec. 3, the experimental results are shown, and conclusions of the paper are given in Sec. 4.

2 The Proposed Superresolution Technique

Figure 1 shows the flow chart of the proposed superresolution technique. Image frames of a video sequence or images from different viewpoints are utilized for the superresolution reconstruction, where a template can be defined to enhance a particular region in the image. The scale-invariant feature transform (SIFT) is adopted to find the corresponding points between the template and the input images. The SIFT extracts image features from images with different viewpoints, and the SIFT matching process discriminates the correct matches from the incorrect matches.[6] Even if the SIFT is robust to variation of scale and rotation, some incorrect corresponding points may exist. Homography does not work well if any incorrect corresponding points are included in the estimation of the homography matrix. Hence, RANSAC is used to robustly match the features from the template and input image, where the correct matches from the SIFT matching process are classified into inliers and outliers.[4,7] Using the inliers selected from RANSAC, we compute the homography matrix **H**, which is used to create isoplane images. The robust RANSAC and weighted homography estimation are proposed for more accurate isoplane transformation. Then, a phase correlation method is applied to the isoplane images for subpixel registration.[8,9] The superresolution reconstruction requires

0091-3286/2008/$25.00 © 2008 SPIE

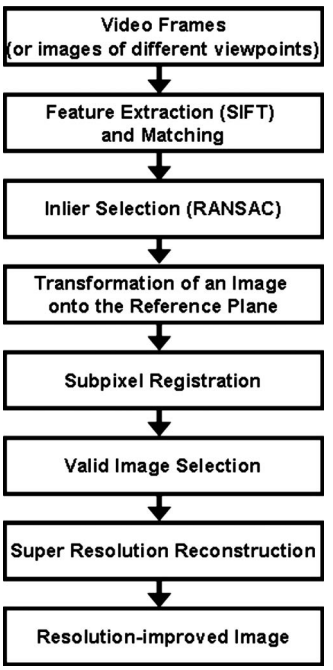

Fig. 1 Flow chart of the proposed superresolution technique.

adequately shifted low-resolution images with subpixel precision. Among many image frames from a video sequence, the images that are the closest to the optimal shifts are selected as valid images. Details of each procedure are described in the following subsections.

2.1 *SIFT and Matching Criteria*

Features from two images are independently extracted using the SIFT, and each feature has a normalized descriptor vector. The SIFT descriptor is created by computing the gradient magnitude and orientation at each sample point in a region around the feature location.[6] A feature region is rotated to be zero-oriented for rotational invariance, and the feature region centered at the feature point is determined in proportion to the feature's scale for scale invariance. After that, the gradient orientation of each pixel in the feature region contributes to the feature descriptor, which has a dimension of $4\times4\times8=128$.[6] The matching parameter between two features is defined as the dot product of their normalized descriptor vectors. As a matching parameter approaches 1, the property of the two features is more similar. Each feature in an image is compared with all features of the other image, and the matching parameters are computed for every comparison. The value of the largest matching parameter is important in deciding the optimum matching between features. In addition, the ratio of the second-largest matching parameter to the largest matching parameter is also an important parameter for optimum matching.[6] If that parameter ratio has a large value (close to 1), there is a rival corresponding point for the matching. If the parameter ratio has a small value, the feature with the largest matching parameter becomes the superior candidate.

In order to investigate the characteristics of the parameter ratio and the largest matching parameter, 10,000 features were obtained from 100 different images, and each feature point had a known corresponding feature point in a different image. In Fig. 2, the horizontal axis is the parameter ratio and the vertical axis is the largest matching parameter. The correct matches from the SIFT matching are mostly at the top left of the plot, and the incorrect matches are near the bottom right. The classification of the correct and incorrect matches in Fig. 2 is a 2-D problem, which is relatively complex. We adopt the principal-component analysis (PCA) method to simplify the 2-D classification to a 1-D classification problem as follows. Using the training features of 10,000 2-D data (the parameter ratio and the largest matching parameter), a covariance matrix can be computed. Since the feature dimension is 2-D, the covariance matrix has a dimension of 2×2. The eigenvector with the largest eigenvalue of the covariance matrix is along the principal axis of the data set.[10] The projection of each datum onto the principal axis is the principal-component value. Therefore, the two quantities—the parameter ratio and the largest matching parameter—for each feature can be converted to one principal-component value, which is called the feature factor p.

The feature factor is a criterion to classify the correct and incorrect matches. In this paper, the classification threshold value of p is 0.2, so that the classification includes 80% of the correct matches and less than 15% of the incorrect matches.

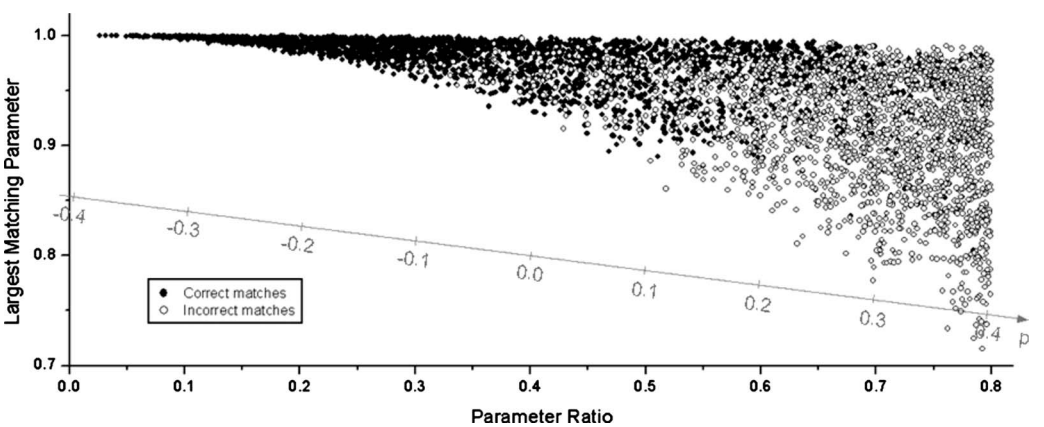

Fig. 2 Distribution of the parameter ratio and the largest matching parameter.

2.2 *Robust RANSAC*

The correct matching features from the SIFT matching[6] are classified into inliers and outliers using RANSAC. Details of the RANSAC process are as follows. First, four sets of corresponding points are randomly selected, and the homography matrix **H** is estimated using those points as follows[7]:

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{1}$$

In Eq. (1), the homogeneous representations of points

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{and} \quad \mathbf{x}' = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix}$$

correspond to $(x,y)=(x_1/x_3,x_2/x_3)$ and $(x',y')=(x_1'/x_3',x_2'/x_3')$, respectively, in Euclidean 2-D space $\mathbf{R}^2$. Since only the ratio of the matrix elements is significant in the homogeneous representation of a point, the homography has eight degrees of freedom. Often, h_{33} and x_3 are set to 1. Equation (1) can be written simply in matrix-vector form as $\mathbf{x}'=\mathbf{H}\mathbf{x}$, where $\mathbf{x}$ and $\mathbf{x}'$ are the corresponding points of the correct matching features from the SIFT matching, and

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

is the homography matrix.

Second, using the homography matrix, the symmetric transfer error $d(\mathbf{x},\mathbf{H}^{-1}\mathbf{x}')^2+d(\mathbf{x}',\mathbf{H}\mathbf{x})^2$ is calculated for every correspondence between two images, and the inliers that are less than the threshold value are counted. Here $d(\mathbf{x},\mathbf{y})$ is the Euclidean distance between points $\mathbf{x}$ and $\mathbf{y}$. Then, the procedures described are repeated for other random selections of four sets of corresponding points. Finally, the case with the largest number of inliers is chosen, where all inliers are used to estimate the homography matrix **H** as explained in Sec. 2.3.

To select inliers, only the geometric distance was used as a criterion in the previous RANSAC. The proposed method considers the geometric distance and the confidence level of the matching features in the RANSAC criterion to determine more robust inliers, as follows:

$$t > C(p)^{-1} \times [d(\mathbf{x},\mathbf{H}^{-1}\mathbf{x}')^2 + d(\mathbf{x}',\mathbf{H}\mathbf{x})^2]. \tag{2}$$

The threshold t is set according to the measurement error, which in this paper is the Euclidean distance between the two corresponding points. If it is assumed that the measurement error has a Gaussian distribution with mean zero and standard deviation σ, then the symmetric transfer error follows a χ^2 distribution.[7] The cumulative χ^2 distribution, $F(k^2)=\int_0^{k2}\chi^2(\xi)\,d\xi$, represents the probability that the value of the χ^2 random variable is less than k^2. From the cumulative distribution, t is defined as $t^2=F^{-1}(\alpha)\sigma^2$, where α depends on the sensitivity level.[7] In this paper, α is set to 0.95, which means that an inlier will be incorrectly rejected with less than 5% probability. Finally, the threshold value is set as $t=(5.99)^{1/2}\sigma$, because $F^{-1}(0.95)=5.99$.[7]

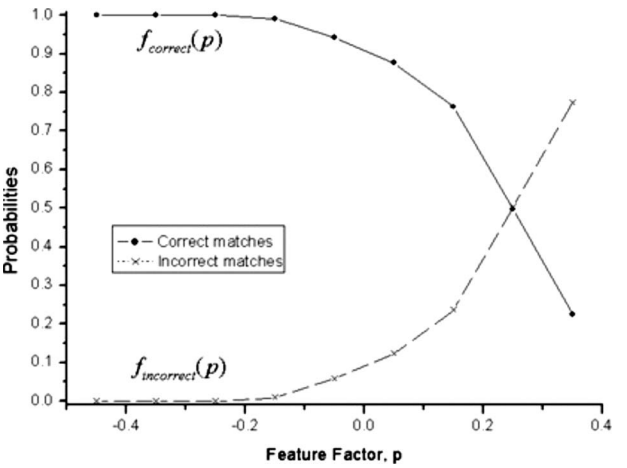

Fig. 3 Probability density functions of the correct and incorrect matches.

In Eq. (2), $C(p)$ is the confidence level of the matching features at $\mathbf{x}$ and $\mathbf{x}'$ as defined in this paper. Here $C(p)$ plays the role of a weighting constant that controls the reward-penalty balance according to confidence in the image description. Equation (2) is efficient in selecting robust inliers, because it considers not only the geometric distance, but also the confidence level of the matching features.

The confidence level of $C(p)$, where p is the value of the feature factor, is inversely proportional to the probability of incorrect matches, so that highly probable incorrect matches can be classified as outliers even though they have a small geometric distance. Similarly, $C(p)$ is proportional to the probability of the correct matches. Therefore, $C(p)$ is defined as

$$C(p) = \frac{f_{\text{correct}}(p)}{f_{\text{incorrect}}(p)}, \tag{3}$$

where $f_{\text{correct}}(p)$ and $f_{\text{incorrect}}(p)$ are the probabilities of the correct and incorrect matches obtained from 10,000 training features, as shown in Fig. 3. In order to model a continuous function $C(p)$ from the measured value $f_{\text{correct}}(p)/f_{\text{incorrect}}(p)$, we used the exponential fitting, which is represented as follows:

$$C(p) = a\exp(-p/b), \tag{4}$$

where $a=7.279$ and $b=0.056$.

2.3 *Weighted Homography Estimation*

Previous workers generally considered that all inliers from RANSAC contributed equally to the estimation of the matrix **H**. However, the matching accuracy may be different for each inlier. This paper proposes a weighted homography estimation to consider the matching accuracy of each

inlier. The homography representation $\mathbf{x}' = \mathbf{H}\mathbf{x}$ in Sec. 2.2 can be rewritten in a different form for all inliers as follows:

$$\begin{bmatrix} x_1^i & x_2^i & x_3^i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1^i & x_2^i & x_3^i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_1^i & x_2^i & x_3^i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} x_1^{i\prime} \\ x_2^{i\prime} \\ x_3^{i\prime} \\ \vdots \end{bmatrix},$$

$$i = 1,2,\ldots,n, \tag{5}$$

or more briefly,

$$\mathbf{A}\mathbf{h} = \mathbf{b}, \tag{6}$$

where $\mathbf{A}$ is a $3n \times 9$ matrix, the 9×1 vector $\mathbf{h}$ is the vector form of the homography matrix $\mathbf{H}$, and $\mathbf{b}$ is a $3n \times 1$ vector. Here n is the number of inliers from the robust RANSAC. In order to estimate h, the weighted least-squares estimation is performed, which minimizes the sum of the squared errors $(\mathbf{b}-\mathbf{A}\mathbf{h})^{\mathrm{T}}\mathbf{W}(\mathbf{b}-\mathbf{A}\mathbf{h})$, where $\mathbf{W}$ is a $3 \times 3n$ diagonal matrix. The diagonal component w_{ii} of $\mathbf{W}$ is determined by the confidence level $C(p)$, where p is the feature factor of the matching features of $\mathbf{x}'$ and $\mathbf{x}^{i\prime}$.

3 Experimental Results

The proposed superresolution technique was applied to real video sequences that simulate surveillance systems in order to double the spatial resolution in both the horizontal and vertical directions. Multiviewpoint images were also used to verify the proposed method.

3.1 *Superresolution of Multiviewpoint Images*

Images with different viewpoints were obtained, and a total of 15 images were captured, each of which had a different viewpoint. The image resolution was 640×480 pixels, and a template of 128×128 was manually selected from an image. Figure 4 shows four images selected from the 15 input images and the template. These input images were transformed by using the robust RANSAC and weighted homography. Subpixel registration was performed, and valid images were selected from the 15 transformed images as explained in Sec. 2. Then, the superresolution reconstruction was applied to the chosen images. For a comparison study, the bilinear interpolation and cubic spline interpolation methods were also applied to the image. The interpolated images are shown in Fig. 5. Text in the low-resolution image could not be recognized well. The interpolation results were better than the low-resolution image. However, some text remained difficult to recognize. In the results of the previous super-resolution reconstruction in Fig. 5(c), most text is enhanced. The result of the proposed method is shown in Fig. 5(d).

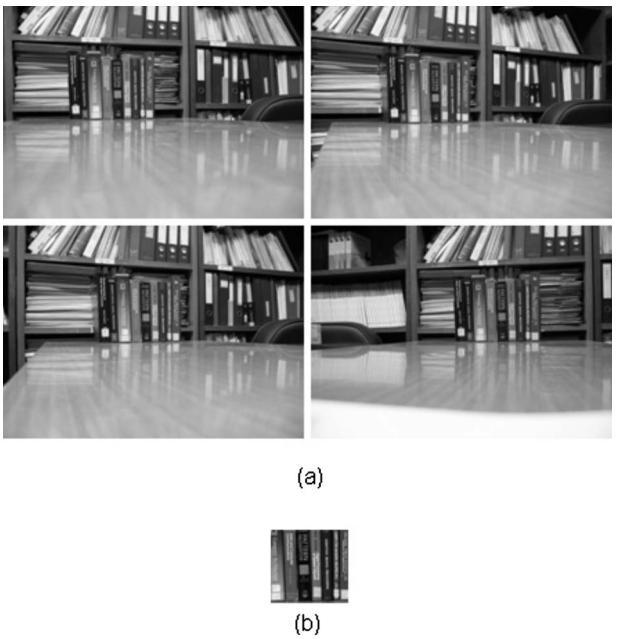

Fig. 4 Four selected images from the 15 images, and the template of the multiviewpoint images: (a) input image frames and (b) the template.

3.2 *Superresolution of Video Sequences with a Moving Object*

A video sequence was obtained from a video camera where the camera position was fixed and a car moved. The image

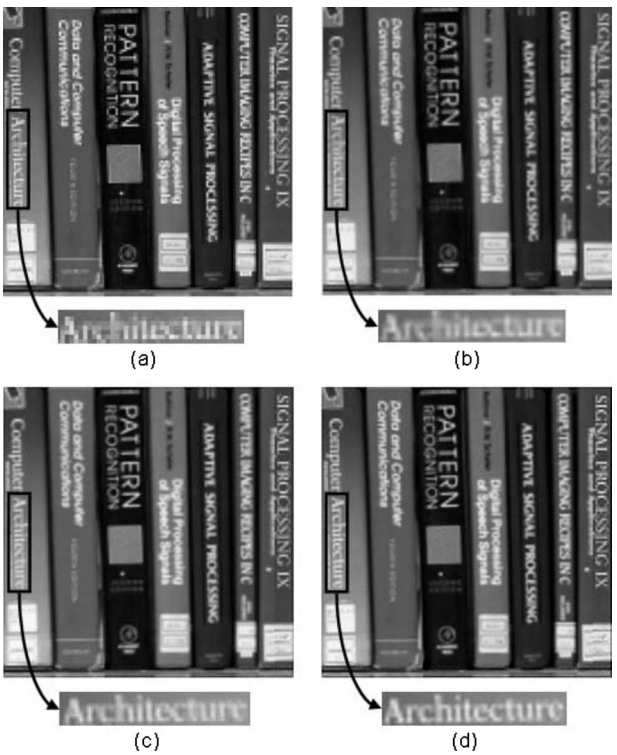

Fig. 5 Resolution-enhanced images from the multiviewpoint low-resolution images: (a) the reference low-resolution image, (b) bilinear interpolation of (a), (c) the previous superresolution reconstruction, and (d) the proposed superresolution reconstruction.

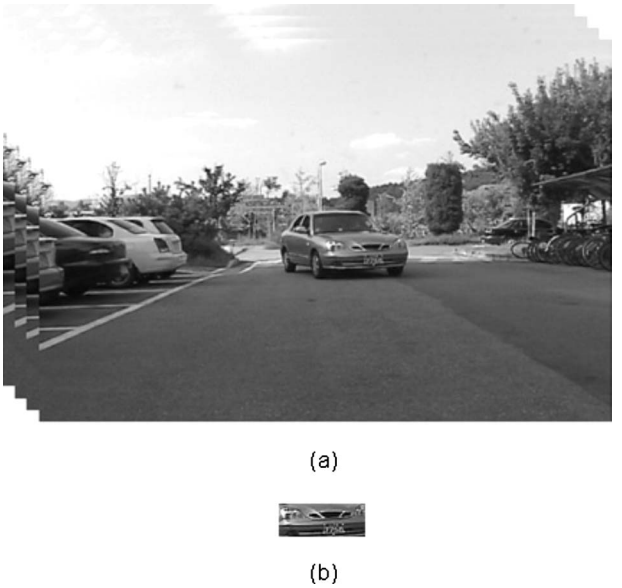

Fig. 6 Image frames and the template of a video sequence: (a) input image frames and (b) the template.

resolution was 720×480 pixels for the 60 input frames, and a template of 104×40 pixels was selected from the input image, as shown in Fig. 6.

Figure 7(a) and 7(b) are two low-resolution images, and Fig. 7(c) and 7(d) are their bicubic interpolation results, respectively. Figure 7(e) is the result from the previous superresolution method where some images were inaccurately transformed. The result from the proposed method is much clearer, as shown in Fig. 7(f). In Fig. 7(a) to 7(d), the numbers in the license plate are not legible. In Fig. 7(f), however, the license plate number can be seen clearly as 7758. The bottom of each part shows the enlarged license plate image.

Fig. 7 Resolution-enhanced images from the video sequence and their license plate regions: (a) a low-resolution image, (b) another low-resolution image, (c) bicubic interpolation of (a), (d) bicubic interpolation of (b), (e) the previous superresolution reconstruction, and (f) the proposed superresolution reconstruction.

In the experiment, the template region was relatively small and blurred. This led to incorrect matches of the features in the previous method. Therefore, the improvements of the proposed method were more distinctive than for the previous experiment in Sec. 3.1.

4 Conclusions

We combined the superresolution algorithm and the geometric transformation in order to deal with multiviewpoint images. For accurate registration between images, we proposed the robust RANSAC criterion and the weighted homography estimation. The robust RANSAC criterion was more reliable than the previous RANSAC criterion because not only the geometric distance but also the confidence level of the matching features was considered. We performed a weighted homography estimation to ensure that more reliable inliers had high weighting values for estimation. The homography matrix, which was obtained from the robust RANSAC and weighted homography estimation, was more accurate and reduced the distortion of isoplane transformation. In conclusion, we confirmed that low-resolution images that were difficult to discern were turned into recognizable high-resolution images using the proposed superresolution technique.

References

1. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Process. Mag.* **20**(3), 21–36 (2003).
2. S. Chaudhuri, *Super-Resolution Imaging*, Kluwer Academic Publishers (2001).
3. E. S. Lee and M. G. Kang, "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," *IEEE Trans. Image Process.* **12**(7), 826–837 (2003).
4. D. Capel and A. Zisserman, "Computer vision applied to super super-resolution," *IEEE Signal Process. Mag.* **20**(3), 75–86 (2003).
5. G. Caner, A. M. Tekalp, and W. Heinzelman, "Super resolution recovery for multi-camera surveillance imaging," in *Int. Conf. on Multimedia and Expo, 2003*, Vol. **1**, pp. 109–112 (2003).
6. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).
7. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge Univ. Press, Cambridge, UK (2003).
8. B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vis. Comput.* **24**, 977–1000 (2003).
9. H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.* **11**(3), 188–200 (2002).
10. http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

Yeol-Min Seong received his BS degree in electrical engineering and computer science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2005. He is currently pursuing his PhD degree in the Department of Electrical Engineering (KAIST). His research interests include image processing, superresolution image reconstruction, and image restoration.

HyunWook Park received his BS degree in electrical engineering from Seoul National University, Seoul, Korea, in 1981, and his MS and PhD degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Seoul, Korea, in 1983 and 1988, respectively. He has been a professor in the Electrical Engineering Department since 1993. He was a research associate at the University of Washington from 1989 to 1992 and was a senior executive researcher at Samsung Electronics Co. Ltd. from 1992 to 1993. He is a senior member of the IEEE and a member of SPIE. He has served as associate editor for the *International Journal of Imaging Systems and Technology*. His current research interests include image computing system, image compression, medical imaging, and multimedia systems.