Correspondence

An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration

B. Srinivasa Reddy and B. N. Chatterji

Abstract— This correspondence discusses an extension of the well-known phase correlation technique to cover translation, rotation, and scaling. Fourier scaling properties and Fourier rotational properties are used to find scale and rotational movement. The phase correlation technique determines the translational movement. This method shows excellent robustness against random noise.

I. INTRODUCTION

Image registration is a fundamental task in image processing to overlay two or more images used. Registration methods can be loosely divided into the following classes: algorithms that use image pixel values directly, e.g., correlation methods [1]; algorithms that use the frequency domain, e.g., fast Fourier transform-based (FFT-based) methods [2]–[12]; algorithms that use low-level features such as edges and corners, e.g., feature-based methods [13]; and algorithms that use high-level features such as identified (parts of) objects, or relations between features, e.g., graph-theoretic methods [13].

The registration method presented here uses the Fourier domain approach to match images that are translated, rotated, and scaled with respect to one another. Translation, rotation, and scale all have their counterpart in the Fourier domain. Fourier methods differ from other registration strategies because they search for the optimal match according to information in the frequency domain [13].

In this correspondence, we present an extension of the phase correlation technique for automatic image registration, which is characterized by its insensitivity to translation, rotation, scaling, and noise as well as by its low computational cost.

This method sometimes relies on the translation property of the Fourier transform, which is referred to as the Fourier shift theorem. Let f_1 and f_2 are the two images that differ only by a displacement (x_0, y_0) i.e.,

$$f_2(x, y) = f_1(x - x_0, y - y_0). \tag{1}$$

Their corresponding Fourier transforms F_1 and F_2 will be related by

$$F_2(\xi, \eta) = e^{-j2\pi(\xi x_0 + \eta y_0)} * F_1(\xi, \eta). \tag{2}$$

The cross-power spectrum of two images f and f' with Fourier transforms F and F' is defined as

$$\frac{F(\xi, \eta)F'^*(\xi, \eta)}{|F(\xi, \eta)F'(\xi, \eta)|} = e^{j2\pi(\xi x_0 + \eta y_0)} \tag{3}$$

where F^* is the complex conjugate of F, the shift theorem guarantees that the phase of the cross-power spectrum is equivalent to the phase difference between the images. By taking inverse Fourier transform of the representation in the frequency domain, we will have a function

that is an impulse; that is, it is approximately zero everywhere except at the displacement that is needed to optimally register the two images.

II. THEORY

In this section, we present the necessary theory to match images which are translated, rotated, and scaled with respect to each other using Fourier translation, rotation, and scaling properties.

A. Without Scale Change

If $f_2(x, y)$ is a translated and rotated replica of $f_1(x, y)$ with translation (x_0, y_0) and rotation θ_0, then

$$f_2(x, y) = f_1(x \cos\theta_0 + y \sin\theta_0 - x_0,$$
$$- x \sin\theta_0 + y \cos\theta_0 - y_0). \tag{4}$$

According to the Fourier translation property and the Fourier rotation property, transforms of f_1 and f_2 are related by

$$F_2(\xi, \eta) = e^{-j2\pi(\xi x_0 + \eta y_0)}$$
$$\times F_1(\xi \cos\theta_0 + \eta \sin\theta_0, -\xi \sin\theta_0 + \eta \cos\theta_0). \tag{5}$$

Let M_1 and M_2 be the magnitudes of F_1 and F_2. Therefore, from (5) we have

$$M_2(\xi, \eta) = M_1(\xi \cos\theta_0 + \eta \sin\theta_0, -\xi \sin\theta_0 + \eta \cos\theta_0). \tag{6}$$

If we consider the magnitudes of F_1 and F_2, then from (6), it is easy to see that the magnitudes of both the spectra are the same, but one is a rotated replica of the other. Rotational movement without translation can be deduced in a similar manner using the phase correlation by representing the rotation as a translational displacement with polar coordinates. i.e., in polar representation

$$M_1(\rho, \theta) = M_2(\rho, \theta - \theta_0) \tag{7}$$

using phase correlation (Section I) angle θ_0 can be easily found out.

B. With Scale Change

If f_1 is scaled replica of f_2 with scale factors (a, b) for the horizontal and vertical directions then, according to the Fourier scale property, the Fourier transforms of f_1 and f_2 are related by

$$F_2(\xi, \eta) = \frac{1}{|ab|} F_1(\xi/a, \eta/b). \tag{8}$$

By converting the axes to lagarithmic scale (this concept is given in [8]), scaling can be reduced to a translational movement (ignoring the multiplication factor $1/(ab)$), i.e.,

$$F_2(\log\xi, \log\eta) = F_1(\log\xi - \log a, \log\eta - \log b) \tag{9}$$

i.e.,

$$F_2(x, y) = F_1(x - c, y - d) \tag{10}$$

where $x = \log\xi$ $y = \log\eta$

$$c = \log a \quad d = \log b. \tag{11}$$

The translation (c, d) can be found by the phase correlation technique and the scaling (a, b) can be computed from the translation (c, d) as

$$a = e^c \quad \text{and} \quad b = e^d \tag{12}$$

where e is the natural logarithm base. If (x, y) is scaled to $(x/a, y/a)$, their polar representation will be

$$\rho_1 = (x^2 + y^2)^{1/2} \tag{13}$$

$$\ominus_1 = \tan^{-1}(y/x) \tag{14}$$

$$\rho_2 = ((x/a)^2 + (y/a)^2)^{1/2}$$
$$= (1/a)(x^2 + y^2) = \rho_1/a \tag{15}$$

$$\ominus_2 = \tan^1((x/a)/(y/a))$$
$$= \tan^1(x/y) = \ominus_1 \tag{16}$$

i.e., if f_1 is translated, rotated, and scaled replica of f_2 their Fourier magnitude spectra in polar representation are related by

$$M_1(\rho, \ominus) = M_2(\rho/a, \ominus - \ominus_0) \tag{17}$$

$$M_1(\log \rho, \ominus) = M_2(\log \rho - \log a, \ominus - \ominus_0) \tag{18}$$

i.e.,

$$M_1(\xi, \ominus) = M_2(\xi - d, \ominus - \ominus_0) \tag{19}$$

where

$$\xi = \log \rho$$
$$d = \log a. \tag{20}$$

Using the expressions (19) and (20) and the phase correlation technique, scale a and angle $\ominus_0$ can be found out. Once the scale and angle information are obtained, the image with the highest resolution is scaled and rotated by amounts a and $\ominus_0$, respectively, and the amount of translational movement is found out using phase correlation technique.

III. IMPLEMENTATION

Before presenting the complete matching algorithm, some important implementation issues are discussed in this section.

A. Cartesian to Polar Conversion

Fourier log-magnitude spectra [8] is used instead of Fourier magnitude spectra for the log-polar conversion. Only the upper two quadrants of the Fourier log-magnitude spectra of images are used to map them to polar plane. This is because of the fact that the Fourier spectrrum is conjugate symmetric for real sequences, as follows:

$$\text{i.e.,} \quad F(\xi, \eta) = F^*(-\xi, -\eta) \tag{21}$$

$$\text{i.e.,} \quad |F(\xi, \eta)| = |F(-\xi, -\eta)|. \tag{22}$$

B. Highpass Filtering

The Fourier log-magnitude spectra are mapped to log-polar plane after multiplying them with a highpass emphasis filter transfer function. A simple highpass emphasis filter is used with transfer function

$$H(\xi, \eta) = (1.0 - X(\xi, \eta)) * (2.0 - X(\xi, \eta)) \tag{23}$$

where

$$X(\xi, \eta) = [\cos(\pi\xi)\cos(\pi\eta)] \quad \text{and} \quad -0.5 \le \xi, \quad \eta \le 0.5. \tag{24}$$

C. Logarithmic Conversion

Logarithmic conversion of ρ-axis is done with base 1.044. This base value is chosen because $\log_{1.044}^{256} = 128$, i.e., 256 rows will be
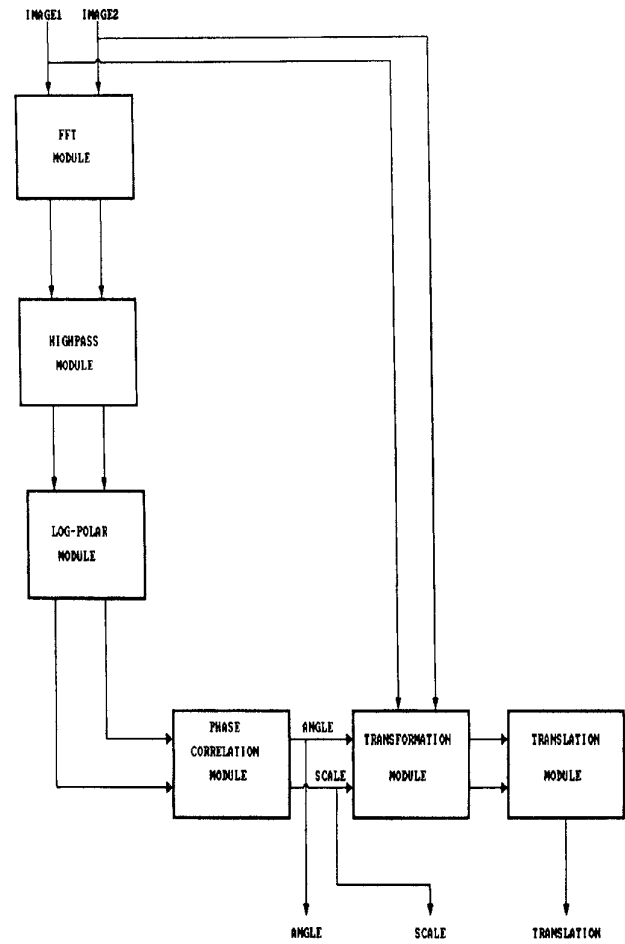

Fig. 1. Overview of the registration algorithm.

mapped to 128 rows in polar plane. Other values of base can also be used depending upon the required level of accuracy.

D. Image Transform

The image transform is implemented using bilinear interpolation from the original image.

E. Fourier Spectra Computation

Fourier spectra of both images can be computed using a single 2-D FFT by taking advantage of the fact that each image data matrix has only real values. This can be achieved by representing one image as the real part and the other image as imaginary part of a complex sequence. Their transforms can be separated from the transform of the complex sequence by "even–odd" separation.

F. Other Issues

We are considering images of size 256×256 and in the log-polar plane representation a size of 256×256. The conversion formulae needed for angle of rotation and scale are

$$\text{scale} = (1.044)^{1/2} \tag{25}$$

$$\text{angle} = (180 * y)/256 \tag{26}$$

where (x, y) is the location of the peak of the inverse Fourier transform of the cross-power spectrum phase. A transformed image
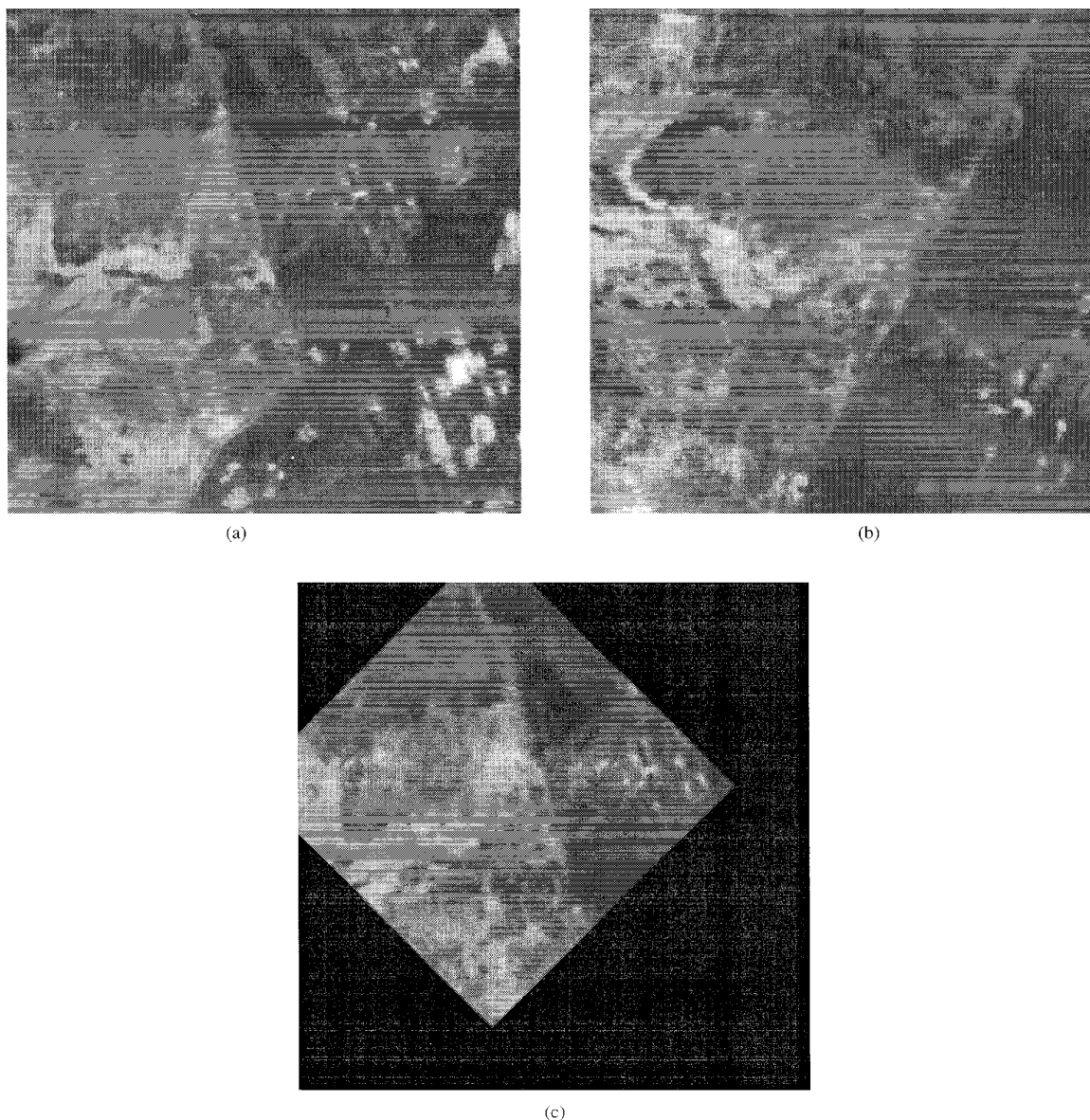
(a)

(b)

(c)

Fig. 2. Registration of translated, rotated, and scaled aerial images.

is obtained from the highest resolution image by using estimated scale and angle information as affine transformation parameters. The transformed image and low-resolution image are used to find the translation without requiring *a priori* information about the image resolution. From several experiments with a set of aerial images, we found out that the matching algorithm can work for scale changes up to 1.8. Corresponding equivalent amount of translation in log-polar domain will be 28. It is likely that the nonlinear complex operations have this limitation. Depending on the order in which the images are supplied to the algorithm, translation will be either x or $256\text{-}x$, on the basis of which image has the highest resolution (this can be easily found out). In determining the angle there is a $180°$ ambiguity. This can be resolved in the following manner. Let θ_0 be the computed angle. First, the translation is determined by rotating the spectrum of one of the images, say Image 2 by θ_0. We then rotate the spectrum of Image 2 by $(180° + \theta_0)$ and again compute the translation. If the value of the peak of the IFFT of the cross-power spectrum phase is

greater when the angle is θ_0, then the true angle of rotation is θ_0, otherwise $(180° + \theta_0)$ is the true angle of rotation.

IV. EXPERIMENTAL RESULTS AND COMPLEXITY ANALYSIS

A. Experiments

The performance of the matching algorithm developed is tested on different sets of images. Different amounts of noise are also added to test the algorithm's robustness against noise. The following types of images are used: i) Images (a) and (b) of Fig. 2 and Fig. 3 of size 256×256 obtained from an aerial image of size 512×512 by applying some known transformations; and ii) cat images (Fig. 4) (size 256×256). Matches are considered valid only if the peak value of the IFFT of the phase difference is greater than 0.03. Theoretically, for exact matches this value should be equal to 1.0; however, the presence of dissimilar parts and the noise in
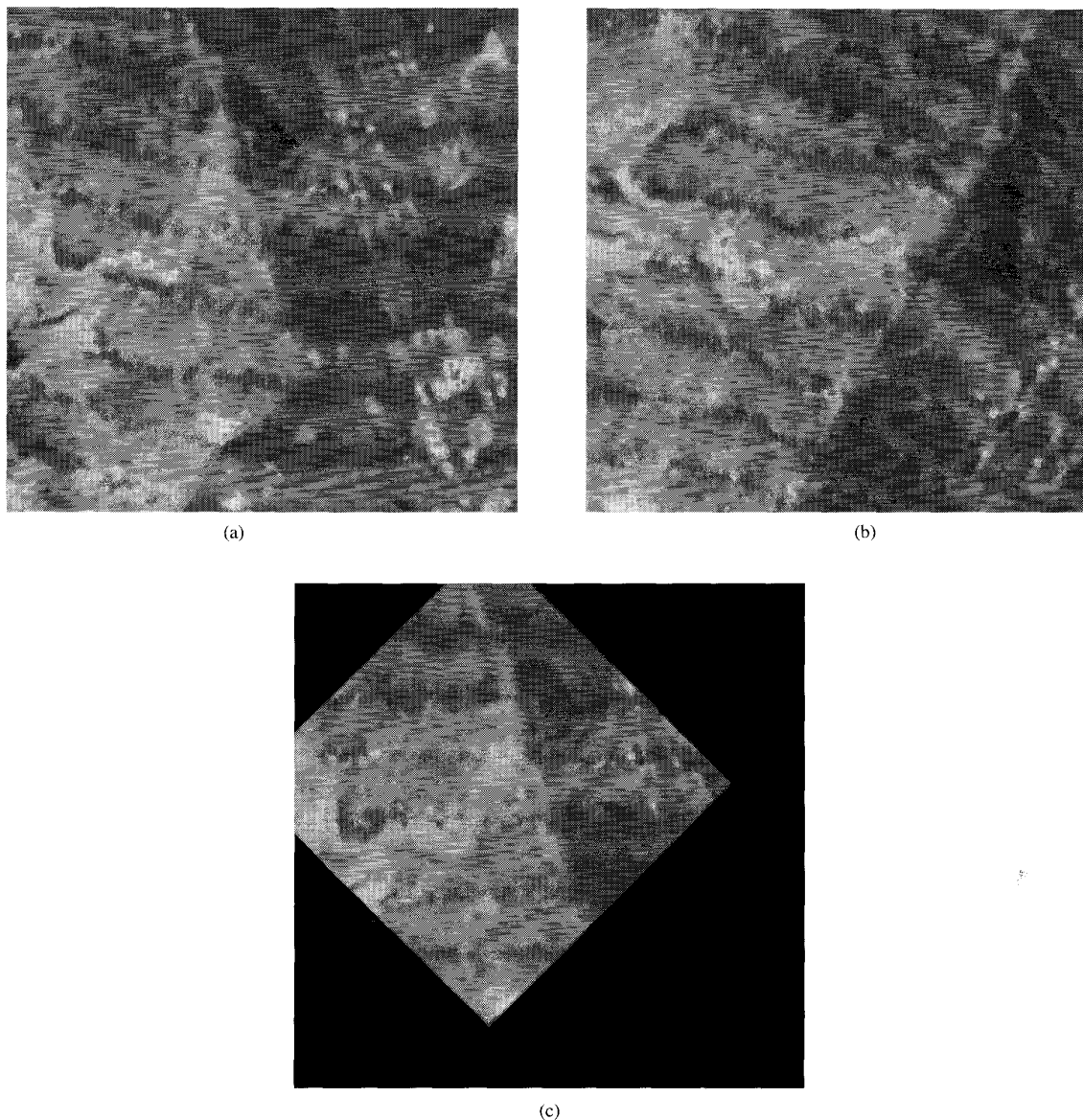
(a)

(b)

(c)

Fig. 3. Registration of translated, rotated, and scaled noisy aerial images.

images reduce the peak value. It has been observed from several experiments that if the peak value is less than 0.03, the match would become unreliable. Results are shown in the following manner: (a) and (b) are test images; (c) shows the resultant image obtained by applying affine transformation on highest resolution image with the parameters computed by the matching algorithm. That is, it shows the matched portion of the images; or, in other words, it shows the exact location of Image 2(b) in the Image 1(a) (assuming Image 2 has the highest resolution). Tables I–III summarize the results. Column "original" shows the exact transformation parameters, column "estimated" shows the computed transformation parameters using the new matching algorithm, "rspeak" indicates the peak value, whose location gives the amount of rotation and scale in log-polar plane, and "tpeak" indicates the peak value, whose location gives the amount of translation. In Tables I and II, matching parameters are listed for the images shown in Figs. 2 and 3. Figs. 2(c) and 3(c) show the registration of translated, rotated, and scaled images, without and

with added noise. Fig. 4 shows the Cat image registration. Here Cat 2 is the highest resolution image; exact transformation parameters are unknown. However, by observing some points in (a) and in (c) it was found (Table III) that the estimated transformation parameter values were equal to the real parameter values, which were obtained manually by actual matching.

B. Complexity Analysis

The proposed matching method requires computation of three FFT's (one to compute Fourier spectra for both images, one to compute Fourier spectra for log-polar converted magnitude spectra, and one to compute Fourier spectrum for the transformed image) and three IFFT's (one to compute rotation and scale, one to find translation, and one to resolve the $180°$ ambiguity in the angle of rotation). On a SUN SPARC Station 1, the proposed method has taken 82 seconds to compute matching parameters for the images of size 256×256. The main advantage of the proposed method is that it
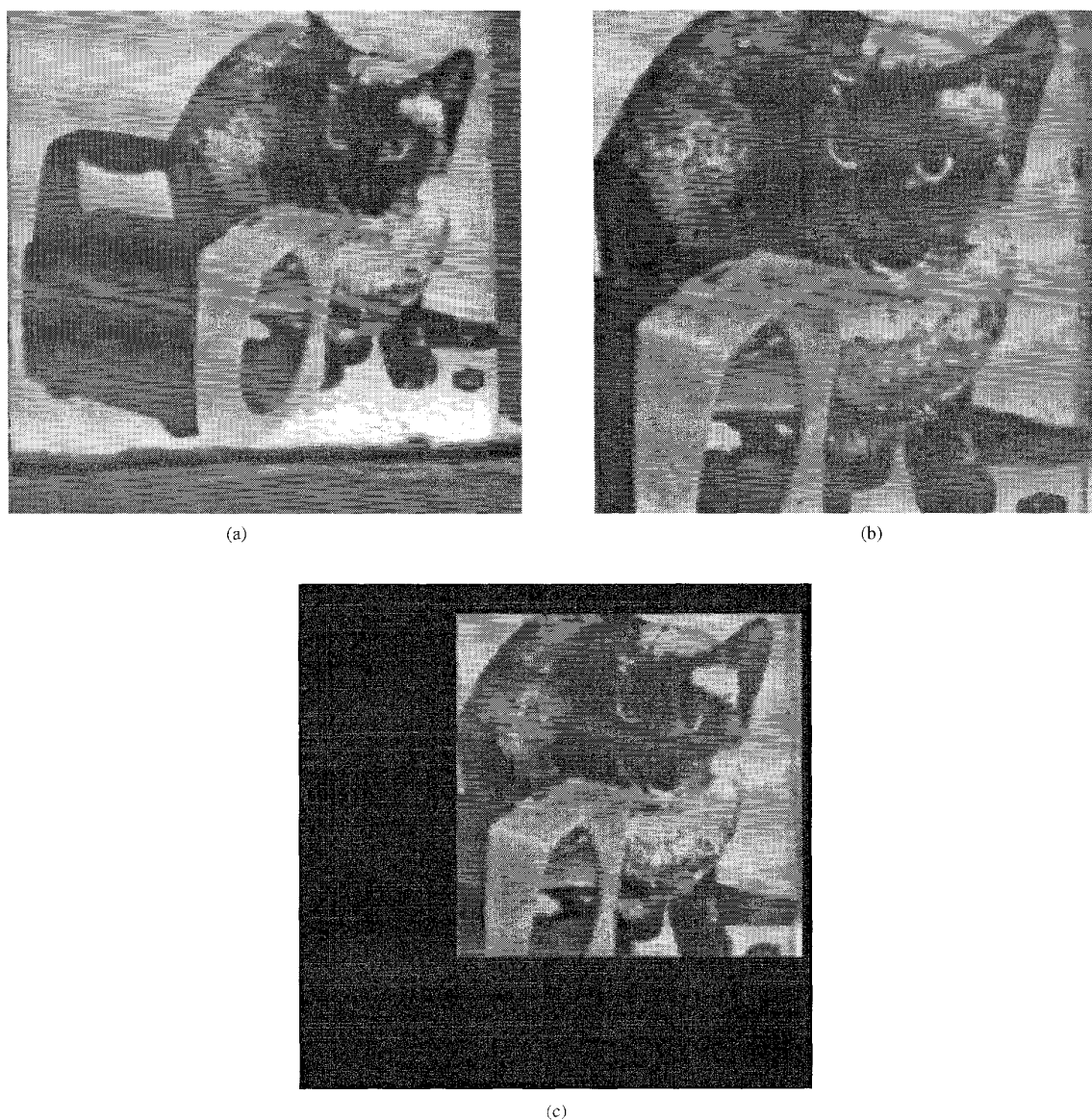
(a)

(b)

(c)

Fig. 4.   Registration of Cat images.

TABLE I
MATCHING RESULTS OF TRANSLATED,
ROTATED, AND SCALED AERIAL IMAGE (FIG. 2)

	original	estimated	
translation	28,30	28,30	rspeak =
rotation	45^0	45.0^0	0.055775
scale	1.47	1.473346	tpeak =
			0.076231

TABLE II
MATCHING RESULTS OF TRANSLATED, ROTATED, AND SCALED
AERIAL IMAGE (FIG. 3); NOISE ADDED TO BOTH IMAGES

	original	estimated	
translation	28,30	28,30	rspeak =
rotation	45^0	45.0^0	0.041178
scale	1.47	1.473346	tpeak =
			0.061093

computes the matching parameters in fixed time for the fixed size of the images irrespective of amounts of translation, rotation, and scale.

De Castro and Morandi's method [5] on an average requires computation of 180 IFFT's. Moreover, for images of size 128 × 128, they have used the spectra of size 512 × 512 to minimize interpolation errors while rotating the spectrum for each angle, and the method fails if scale change is present. Also, it cannot compute matching parameters in constant time.

V. CONCLUSION

In this work, a FFT-based matching, which is an extension to phase correlation technique, was presented. The proposed algorithm

TABLE III
MATCHING RESULTS OF CAT IMAGE (FIG. 4)

	original	estimated	
translation	-,-	-26,37	rspeak =
rotation	-	0.0⁰	0.059670
scale	-	1.473346	tpeak =
			0.204867

is characterized by its insensitivity to translation, rotation, scale, and noise as well as by its low computational cost. For a given size of images, it computes matching parameters in constant time irrespective of the type of images and amounts of translation, rotation, and scale. The matching is found to be exact except for scaling where it is found to deviate in the third place of decimal probably due to the nonlinear processing.

Existing FFT-based matching methods fail if scaling is present between the images. Also, their computational cost is high. All these difficulties are successfully suppressed in the proposed matching method.

In Section I, on the basis of several experiments, we have found out that the matches are valid if the peak value of the IFFT of the phase difference is greater than 0.03. In fact, one can study the effect of the variation in the size of this peak on registration. One can probably use the size of the peak to determine how good the match is.

REFERENCES

[1] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital registration," *IEEE Trans. Comput.*, vol. C-21, pp. 179–186, 1972.

[2] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proc. IEEE 1975 Int. Conf. Cybernet. Society*, New York, NY, pp. 163–165.

[3] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Appl. Opt.*, vol. 23, no. 6, pp. 812–816, 1984.

[4] S. Allney and C. Morandi, "Digital image registration using projections," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 222–233, Mar. 1986.

[5] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-95, pp. 700–703, Sept. 1987.

[6] D. J. Lee, T. F. Kpile, and S. Mitra, "Digital registration techniques for sequential fundus images," in *IEEE Proc. SPIE: Applic. Digital Image Processing X*, New York, 1987, vol. 829, pp. 293–300.

[7] A. Appicella, J. S. Kippenhan, and J. H. Nagel, "Fast multi-modality image matching," *SPIE Med. Imaging III: Image Processing*, vol. 1092, pp. 252–263, 1989.

[8] D. Casasent and D. Psaltis, "Position oriented and scale invariant optical correlation," *Appl. Opt.*, vol. 15, pp. 1793–1799, 1976.

[9] H. Wechsler, *Computational Vision*. New York: Academic, 1990, pp. 105–108.

[10] P. E. Zwicke and I. Kiss, "A new implementation of the Mellin transform and its application to radar classification of ships," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 191–198, 1983.

[11] Y. Sheng and J. Duvernoy, "Circular-Fourier-radial-Mellin descriptors for pattern recognition," *J. Opt. Soc. Amer. A.*, vol. 3, no. 6, pp. 885–887, 1986.

[12] Y. Sheng and H. H. Arsenault, "Experiments on pattern recognition using invariant Fourier-Mellin descriptors," *J. Opt. Soc. Amer. A.*, vol. 3, no. 6, pp. 771–776, 1986.

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

A Comparison of Different Coding Formats for Digital Coding of Video Using MPEG-2

Gisle Bjøntegaard, Karl Olav Lillevold, and Robert Danielsen

Abstract— This correspondence adresses the problem of maximizing the subjective picture quality at given bit rates for coded video. Coding is carried out using different picture formats. The parameters involved are the number of pixels horizontally and vertically and interlacing/noninterlacing. The coding is done according to the MPEG-2 standard.

I. INTRODUCTION

The MPEG-1 and MPEG-2 [1], [2] coding methods for video compression belong to a class of methods referred to as *hybrid discrete cosine transform (DCT)*. The ITU H.261 and H.263 Recommendations also belong to this class. These methods are well documented elsewhere and will therefore not be introduced in detail here. We will focus only on the aspects of special importance for this work.

The MPEG methods are largely based on dividing the picture material into 16×16 blocks of picture material (macroblocks). For each of these blocks there is a prediction block, which is found in previously decoded pictures by using displacement vectors. In some cases the prediction is sufficient to represent the 16×16 pixels. However, most of the time a difference block has to be transmitted. For the 16×16 block, a difference block is produced by subtracting the prediction from the block to be coded. The difference block undergoes a two-dimensional 8×8 DCT. The transform coefficients are then quantized before being entered into a bitstream. The quantization parameter Q determines the level of quantization; we may view this as the "coarseness" of the resulting picture. The parameter Q referred to in this work is the actual divisor used for quantization. To a large extent the size of Q therefore determines the number of bits produced from a 16×16 block. The resulting number of bits can be considered to consist of two parts: i) motion vector data for prediction plus general overhead and ii) transform coefficients.

The transform coefficients usually account for the largest part. The number of bits used for transform coefficients is influenced by the size of Q and may therefore directly be controlled by the rate control mechanism employed. The rate control mechanism uses a buffer fullness measure to give feedback to the adjustment of Q.

In this correspondence, a format or picture format is defined by the horizontal and vertical pixels in a picture as well as the interlaced or progressive structure of pixels (see Figs. 1, 2, and 3). We will talk about interlaced and progressive formats. We will also talk about low-resolution formats and high-resolution formats indicating whether a format has few or many pixels.

The MPEG standard does not give detailed rules of how pictures shall be coded. The standard is, rather, a protocol for the decoding procedure. Another way to see this is that MPEG provides a toolbox that may be used by the encoder to obtain good compression performance. This means that many choices are left for the encoder. The present paper examines the combination of three important