MTA SZTAKI Közlemények 32/1985 123-140

AN EFFICIENT SYNTHESIS OF IMAGE MATCHING ALGORITHMS

HOANG KIEM, PHAM NGOC KHOI
Institute of Informatics and Cybernetics
Hanoi-Vietnam

ABSTRACT

In order to improve the efficienty of image matching, a lot of matching schemes have been proposed, based on various approaches [1-5]. Here we discuss the efficiency of the synthesis of image matching algorithms using hierarchical schemes and those that use the combination of coase-fine matching algorithms. The method of extracting the features for regions in an image and perfomance of scene matching methods are considered.

This paper consists of the following parts:

- On the synthesis of image matching algorithms
- k-centroid feature extraction for image matching
- Combination of image transformation and normalization
- Synthesis scheme for image matching programs

1. ON THE SYNTHESIS OF IMAGE MATCHING ALGORITHMS

It is well known that the problem of scene matching is given a template of a scene, determine the location of this template in another scene. The method used to solve this problem, in its simplest form, is called template matching with the basic correlator, the statistical correlator. Later, some modifications using invariant moments for scene matching have been developed to solve the general problem involving geometrical and sensor variation [4-5].

Since a template of size MxM can be shifted into $(N-M+1)^2$ possible positions in an image of size NxN, the number of coorelations can be extremely large. The tendency in the current research is toward the use of hierarchical techniques for decreasing the number of search position. In particular, coarse-fine techniques are logarithmically efficient and reduce the number of search positions to $K.log(N-M+1)^2$, where K is a constant. [1,2,3].

However, at level of search, the number of computations needed to obtain the features for scene matching for example, invariant moment can be still large. Later, a synthesis using hierarchical technique and detections is proposed.

1.1 Hierarchical schemes for image matching

- At first, a structured set of pictures at different resolution is defined. The level K scene is reduced to a level $(K-1)$ scene with the agglomerative rule, for example:

$$F_{K-1}(i,j) = \frac{1}{4}\{ F_K(2i,2j) + F_K(2i,2j+1) + F_K(2i+1,2j) + F_K(2i+1,2j+1)\}$$

where, $F_K(i,j)$ - the gray scal of pixel (i,j) at level K. Note that, at the level K, number of possible test locations is $[(N-M+1)/(2^K+1)]^2$ and at level $K-1$, only the locations selected in level K needed to be tested.

- A matching rule to guide the search from level $K-1$ to level K must also be defined. In the scene matching with invariant moments, this rule is the moment correlation which is costly in computation, due to the calculations needed to obtain the invariant features. But it can be used to great advantage at the low resolution level at which other methods are not possible. Here, we use an approach as follows: instead of matching each template of scene at every location, the templates are partitioned into "informative" and "irrelevant" templates by some simple tests. Elimination of mismatching locations and termination of computation can take place at each level of test based on this partition.

In practice, we have used a detection that combined two cimple tests before matching the secene with the invariant moments:

1°. Test based on measure of the similarity of two gray level distributions (τ-test).

2°. Test based on the correlation coefficient of the joint distributions (ρ-test). The τ, and ρ measures are computed for each location. If both τ,ρ are smaller than selected threshold, this location is rejected.

- Thus, let N_k^i be a set of test locations (u,v) at search level K, with a matching rule R_k^i such that

$$N_k^i = \{(u,v)\,|\,R_k^i(u,v) \geq \theta_k^i,\ 1 \leq u,v \leq M\}$$

where θ_k^i is the threshold selected to be used at search level K, R_k^i is some matching ruleat test location (u,v), M is the number of picture elements in the template. We can divide R_k^i into the preliminaty rule detection by simple test and the main rule (for example, the moment correlation rule).

Let $N_k := \bigcap_i N_k^i$, for a search region of size NxN, an $(2N-2M+1)^2$-matrix G_{k-1} was generated by

$$G_{k-1}(2i,2j) = \begin{cases} 1 & if \quad (i,j)\in N_k \\ 0 & if \quad (i,j)\bar{\in} N_k \end{cases}$$

All other entries of G_{k-1} are set to zero. Tests are to be performed at the test locations for $G_{k-1}(u,v)=1$. The search continues untill one of two conditions is encountered

1°. At search level $L=n$, $G_n(u,v)=1$ for one value of (u,v), location (u,v) is declared the matched location.

2°. At the level $L=0$, there exist severeal locations (u,v) such that the $G_o(u,v)=1$. Select the location with the higest correlation with the invariant moments.

1.2 Theorem

The condition for savint the computation time using this synthesis is following:

$$\Phi' < \Phi(1-p)$$

where Φ' - the computation complexity of the detection at each location defined by the number of calculation needed.

Φ - the computation complexity of the main-matching rule at each location defined by the number of calculation needed.

p - the probability of matching by the detection.

Proof:

Noting that, at level k, number of possible test locations is $[(N-M+1)/(2^k+1)]^2$ then the number of calculation needed for scene matching using this synthesis is

$$\mathcal{N}_s = [(N-M+1)/(2^k+1)]^2 \cdot \Phi' + [(N-M+1)/(2^k+1)]^2 . \Phi . p$$

For saving the computation time using this synthesis, the following condition need to be satisfied:

$$\mathcal{N}_s < [(N-M+1)/(2^k+1)]^2 \Phi$$

It follows that

$$[(N-M+1)/(2^k+1)]^2 . \Phi' + [(N-M+1)/(2^k+1)]^2 \Phi . p$$
$$< [(N-M+1)/(2^k+1)]^2 \Phi$$

Dividing both sides of the above inequality to $[(N-M+1)/(2^k+1)]^2$ we have

$$\Phi' + \Phi . p < \Phi$$

or $\Phi' < \Phi(1-p)$

which proves the theorem.

Theorical analysis and simulation with τ-test and ρ-test in scene matching by invariant moments indicated that a saving of computation time as well as a high degree of precision in locating a region is possible.

1.3 The τ-test

The χ^2-test measures the difference between two frequency distributions. Let $h_m(k)$ be the frequency distribution of gray-scale intensities in a model window. Let $h_t(k)$ be the frequency distribution of a test window. The significance of the χ^2-test depends on the number of samples:

$$\chi^2 = \sum_k \frac{(h_m(k)-h_t(k))^2}{h_t(k)}$$

where, we can consider h_m to be a hypothetical ideal distribution. Let $\tau = e^{-\chi^2/c}$, where c is some positive constant. τ is a measure of the similarity of two distributions (in the χ^2 sense) If the distributions are identical, then τ will be unity, if they are very different, τ will be close to zero.

τ is not sensitive to the location of pixels. It simply measures the degree of simplarity between two marginal distribution.

1.4 The ρ-test

Let m_1 be the mean of h_m and m_2 be the mean of h_t. Let σ_1 be the standard deviation of h_m and σ_2 be the standard deviation of h_t. We define the coefficient ρ as follows:

$$\rho = \frac{\mu(1,1)}{\rho_1 \cdot \rho_2}$$

where, $\mu(i,j) = \frac{1}{n} \sum_k (x_m(k)-m_1)^i \cdot (x_t(k)-m_2)^j$.

$x_m(k)$ and $x_t(k)$ are

are gray velues of the k-th pixel in the model image and the test image, respectively.

ρ is in the interval $[-1,1]$. In general, if there is a linear functional dependence between the test and model window ρ will be 1. If the window are independent distributions ρ will be 0. Thus, the intermediate values will measure the degree of dependence between the two windows.

ρ is a good test for the proper location of pixels. With the systematic change in lighting, τ would be small but ρ would be large because the test and model distributions would still be well-correlated.

1.5 Performance of scene matching methods

To evaluate the performance of any of matching techniques one may consider the probability distribution that could be at the k-th level in the hierarchical search or the first level for template matching. The distribution $p_k(R)$ is the probability that the true match location takes on a specific similarity value R. $R_k(u^*,v^*)$ is the similarity value at the true math location for a particular match under consideration. Let P_k be the probability of detection of thek-th search level (i.e. the probability that the similarity measure at the true match location exceed the threshold R_T^k) and the $p_k(R)$ will be assumed to have a Gaussian distribution with a variance of $(\sigma_R^k)^2$. Then p_k can be expressed ad

$$p_k = \int_{R_T^k}^{\infty} p_k \; R \; dR = \frac{1}{\sqrt{2\pi}} \int_{y}^{\infty} exp(\frac{-R}{2})^2 \; dR$$

where

$$y = [R_T^k - R_k \; (u^*,v^*)]/\sigma_R^k$$

Similarly, the probability of false fix at the k-th search level can be computed by

$$P_f = \int_{R_T^k}^{\infty} p_f(R)dR = 1-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} exp(\frac{-R}{2})^2 dR = 1 - \Phi(y)$$

where $y = (R_T^k - R_b^k)/\sigma_f^k$, $p_f(R)$ is the probability density distribution of the similarity measures of all test locations except the true match location with a variance of $(\sigma_f^k)^2$, R_b^k is the similarity measure averaged over all test locations (Fig.1)

The error introduced by the Gaussian assumption for a non-Gaussian distribution may be large in the case of small values of $p_k(R)$. We can use the Edgeworth expansion for this case.

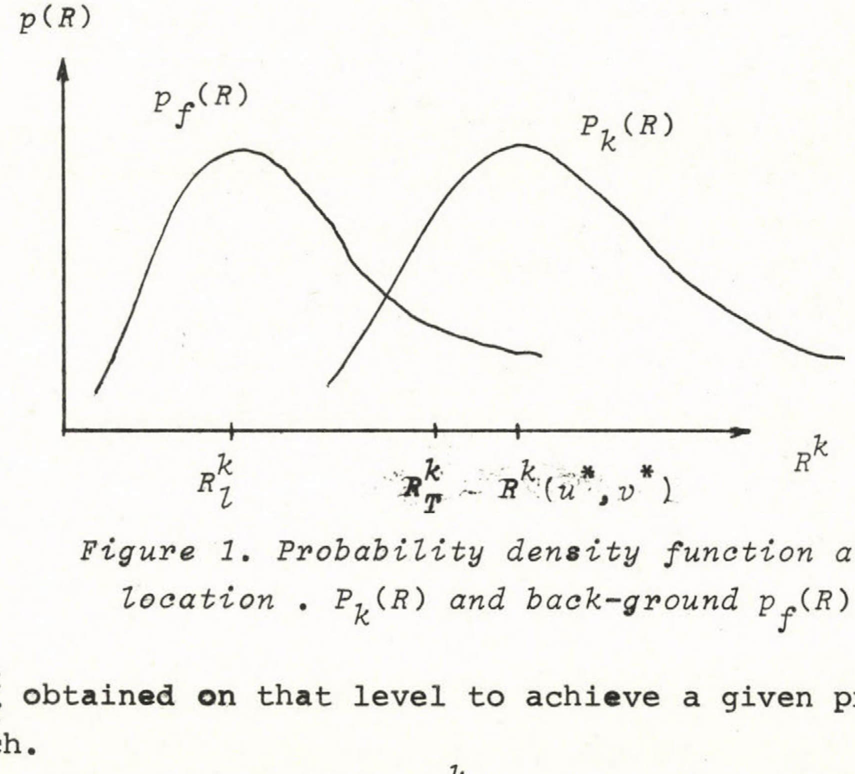

Figure 1. Probability density function at match location . $P_k(R)$ and back-ground $p_f(R)$

σ_R^k obtained on that level to achieve a given probability of a match.

To select the threshold R_T^k, the invariant moments of the image were correlated with the moments of the image reduced by a factor of 2 and rotated by 2° and 45°. The averaged correlation of the three cases was then used as a bound to estimate a threshold sequence. Noting that the correlation were made on the logarithms of amplitude of the moments rather than the amplitude itself.

2. FEATURE EXTRACTION FOR IMAGE MATCHING

At the fine-matching step, the moment approach is often used for this purpose. However, scene match, with invariant moments is costly in computation. In many cases, the following approaches give us the poweful features for image matching with smaller computation time.

2.1 The kG-centroid (kL-centroid) features

We define a kG-centroid (kL-centroid) of the image as a centroid of this image at gray level k (of k-th region of this image). Suppose that (x_c^k, y_c^k), $k=1,2,\ldots K$ are kG-centroid (kL-centroid) and (x_c, y_c) is centroid of the image I.

Let (r^k, θ^k) be polar coordinates of (x_c^k, y_c^k) in the polar coordinate system with (x_c, y_c) as the origin and the rayon passed (x_c, y_c), (x_c^k, y_c^k) as the initial rayon k^o is a chosen value, say, $k^o = 1$.

By converting (r^k, θ^k) into $(\tilde{r}^k, \theta^k)$, where

$$\tilde{r}^k = r^k / \sum_{i=1}^{k} r^i$$

then $(\tilde{r}^k, \theta^k)$, $k=1,2,\ldots K$, are invariant features in relation to translation, rotation and size change.

We can also derive invariant features from kG-centroid (kL-centroid) in the following manner.

Let $\rho(k,k')$ be the distance between (x_c^k, y_c^k) and $(x_c^{k'}, y_c^{k'})$ where $k,k'=1,2,\ldots K$, then $\{\rho(k,k')_{k,k'=1,2,\ldots K}\}$ are dependent only on the shape of the object, but not affected by its location, orientation or relative size.

For the normalized images, we can extract simplier features as follows.

2.2 Projections and Cross-Sections

Given a two-dimensional continous function $f(x,y)$ we define the projection of $f(x,y)$ in the x and y direction are

$$\int_R f(x,y)dx \qquad \text{and} \qquad \int_R f(x,y)dy$$

In principle, projections in a sufficient number of directions contain enough information to reconstruct the picture [5].

For a digital image, the $x(i)$ and $y(j)$ projections are defined as:

$$x(i) = \sum_j f(i,j)$$

$$y(j) = \sum_j f(i,j) \qquad \text{for} \quad 1 \leq i,j \leq N$$

More detailed information about the arrangement of gray levels in the region R can be obtained by using projections of $f(x,y)$ in various directions.

In many cases, the projections on X-Y axes a certain amount of information of object for recognition and the number of dimension may be significantly reduced from N^2 to 2N, (N is the dimension of the image). Furthermore, the numerical properties of projections, such as their (one-dimensional) moments, Fourier coefficients, Walsh coefficients, etc... can be used as the powerful features for recognition of the image in which, most of the strokes of patterns are either horizontal or vertical and they generate many step segments in the projections. Some experiments have been made successfully by these features [13].

3. COMBINATION OF IMAGE TRANSFROMATION AND NORMALIZATION

As we know that, the power spectrum of an image is to be independent of translation. The Mellin transform has been show to be scale independent. The Polcar-Cartesian transform convers rotation into translation. Hence a combination of these performed successively will allow shape to be matched to shape independent of translation, rotation and scale [10,15].

Here we use a simple normalization scheme, the normalized image which also is invariant to object translation, rotation and size change.

- Normalization in relation to translation

The image is normalized to an image-centered coordinate system with its centroid is translated into a fixed point

$$N(x,y) = I(x_o+x-x_c, y_o+y-y_c)$$

where, (x_o, y_o) be a fixed point, (x_c, y_c) be the centroid of image.

- Normalization in relation to rotation

The image is normalized to coordinate system with its principal axes as coordinate axes

$$N(x,y) = I(x, cos\theta - y.sin\theta, x.sin\theta + y.cos\theta)$$

where

$$\theta = \frac{1}{2}\, tan^{-1} \frac{2\, I_{11}}{I_{20}-I_{02}} ,\quad I_{pq} = \sum_{s} (x-x_c)^p (y-y_c)^q\ f(x,y)$$

- Normalization in relation to size change

The image is scaled to a standard size

$$N(x,y) = I(k_x.x, k_y.y)$$

where (k_x, k_y) be the ratio of the size of the image I to standard size.

In this way, the normalized image dependent only on the shape of the object, but not affected by its location, orientation, or relative size.

4. SYNTHETIC SCHEME FOR IMAGE MATCHING PROGRAMS

In connection with scheme of synthesis of image matching algorithms (Fig. 3) we propose an use manner of programs as follows.

Given an image of size NxN and a template of size MxM, if necessary, we make a geometrical correction for input image by using GECOR (See appendix). The next phase is coarse matching the template to windows of image. The propesed step matching is M/2. In this phase, we combine several tests (τ-test, ρ-test) by using HISTO, PROFIL, THRSLD (See appendix). So we obtain the possible match locations.

In the next phase, with each of the possible match locations, we make a fine matching around those locations. For improving the efficiency of fine matching, both the template and the window may be modified either by transforms Polcar, Mellin, Fourier using TPOCAR, MELIN, TFØ (See appendix) or by normalizations: centered translation, rotation, scaling using TCEN, TROTA, SCALEØ (See appendix). With these transformations (normalizations), the transformed (normalized) template and window will be invariant to rotation, translation, scale... After that, we can use the simple features such as projections for matching. Then the fine matching may be made by computing the feature correlation: invariant moments, kL-centroid (kG-centroid), profils...using MOMENT, TOPO, PROFIL, THRSLD (See appendix). We will present program descriptions in the appendix.

5. CONCLUSIONS

Experimental results indicated that scene matching with the basic seguential method provided good performance in the matching of scene that contain relatively weel-defined objects of varying background. This method is particularly useful in matching images taken by the same type of sensors under different operating conditions. Depending on the scene content, scene matching with invariant moments was successful in some cases. In particulary, this method can be used to great advantage at

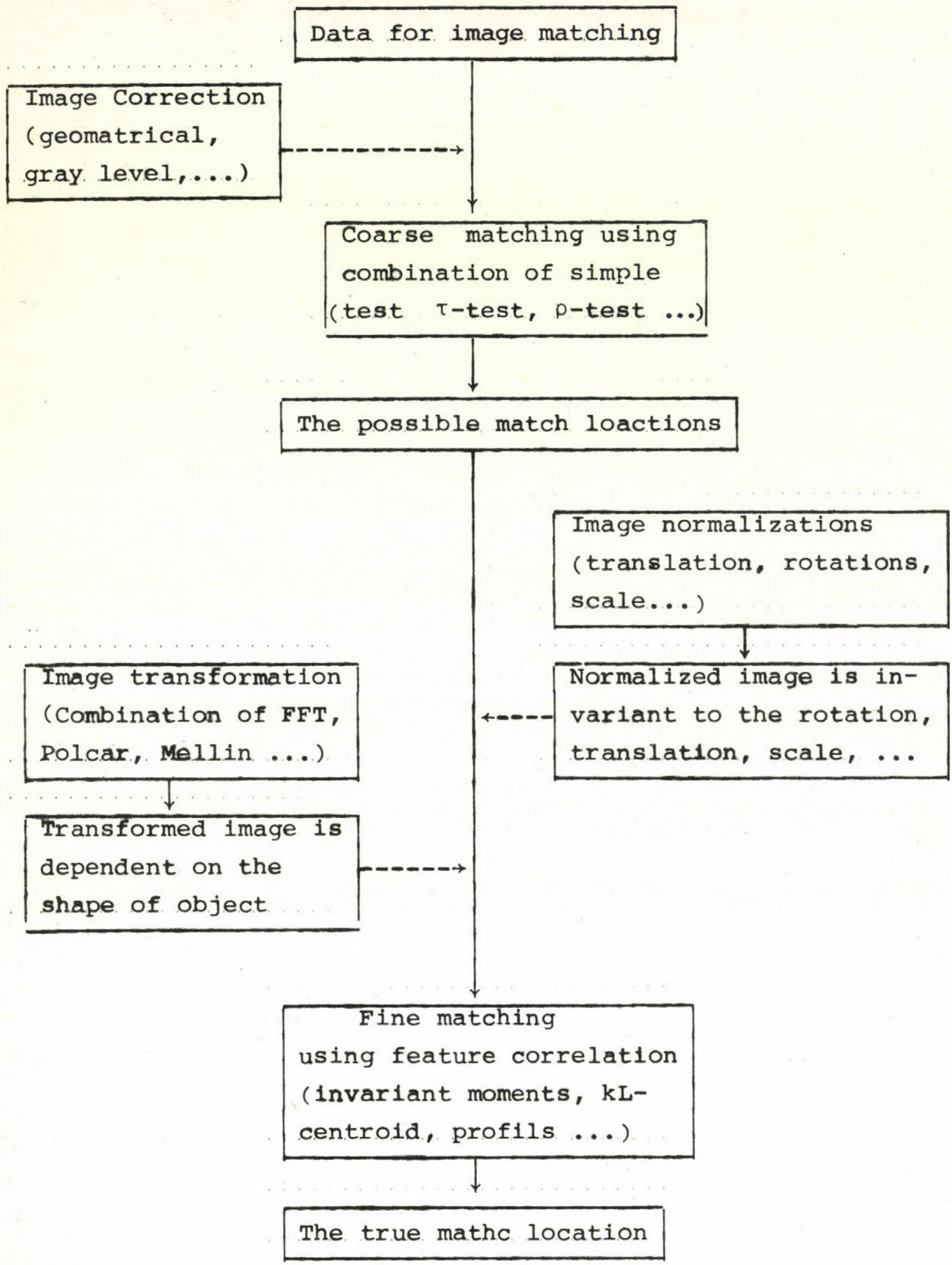

Fig. 3. The scheme of synthesis of image matching algorithms

the low resolution level at which other methods, such as scene matching with edge features are not possible. Two improvements may be accomplished are the following:

- Weight each of the moments with an appropriate weighting factor before correlation.

- Generate higher-order moments. Select a set moments for correlation computation with the selection based on the information contents of the images.

At the same time, we also have used some other approaches as follows:

- The comination of transformations Fourier, Polcar, Mellin
- The image normalization in relation to certain transformations as translation, rotation, scale ...
- The kG-centroid (kL-centroid) features.

These methodes have resulted in superior performance and were accomplished at greatly reduced computation and memory storage requirements.

REFERENCES

1. Y.Wong et al.: Scene matching with invariant moment. CGIP 8/1978, p. 16-24.
2. Y.Wong et al.: Sequential hierarchical scene matching. IEEE Trans on Comp. Vol C-27, 1978, p. 359-366.
3. G.Kanderburg: Coarse-fine template matching. SMC 7/1977. p. 104-107.
4. E.Hall: Computer image processing and recognition. Academic Press. New York 1979.
5. A. Rosenfeld et al.: Digital Image Processing. New York 1980.
6. A.Rosenfeld et al.: Relative effectiveness of selected texture primitive statistics for texture discrimination. SMC-11, 5/1981, p.360-370.
7. M.K.Hu: Visual pattern recognition by moment invariants. IRE Trans inform theory. Vol 67. 2/1962, p. 179-187.
8. S.Maitre: Moment Invariants. Proc. IEEE Vol 67, 4/1979. p. 697-699.
9. T.Hsia: A note on invariant moments in image processing. SMC 12/1981. p. 826-831.
10. N.Ahmed et al.: Orthogonal transforms for digital signal processing. New York 1975.
11. R.N.Bracewll: The Fourier transfromation and its application New York 1965.
12. R.Chinins: A complet set of Fourier Descripter for two-dimensional shapes. SMC-12, 1982, p. 848-855.
13. H.Arakwa: On-line recognition of hadwritten characters. PR. Vol 16, 1/1983, p.9-16.
14. D.Casasent et al.: New optical Transformation for Pattern Recognition. Proc. IEEE Vol 65, 1/1977, p.77-84.
15. R.Michael: Digital image shape detection. AFIF Vol 48 1979, p. 243-254.
16. V.Wedler: An unified approach for invariant properties and normalization of images. Rostock-Math-Kolloq-20, 1982. p.53-60.
17. C.H.Chien et al.: A normalized quadtree representation. CGIP 26, 1984, p. 331-346.

APPENDIX I: A SET OF PROGRAMS FOR IMAGE MATCHING

Programs are written in FORTRAN-4 subroutines and were verified on PDP-11 minicomputer system. In the following subroutines

IMAG is an input image, a variable-array of size NxN

I,J are coordinates of the upper-left corner of the windows

1. PROGRAM FOR GEOMETRICAL CORRECTION OF IMAGE

The call   CALL GECOR(IMAG,V,N,X,Y,XF,YF,K,MOD)

V : output image of size N N

X,Y the given points in IMAG

XF,YF : the given points in IMAG, which correspond to X,Y in V

K : length of X,Y,XF,YF

MOD : correction mode

MOD=0 correction by the way of 4-neighbor average

MOD=1 correction by the way of nearst neighbor

2. PROGRAMS FOR COARSE MATCHING

2.1 Computation of Histogram of a window

The call   CALL HISTO(IMAG,I,J,K,H,LEVEL)

K : size of window

LEVEL : gray level of image

H : output histogram of size LEVEL, in which H(i) is number of pixels of gray level i+1.

2.2 Computation of Profil of a window

The call   CALL PROFIL(IMAG,I,J,K,HX,HY)

K : size of windos

HX,HY : vectors of size K- profils on the x-axe and y-axe

2.3 Thresholding program

The call   CALL THRSLD(A,K,THR)

A : input data set of size K to be thresholded

THR : output threshold

3. PROGRAMS FOR NORMALIZATION OF IMAGE

3.1 Determination of the image center

The call   CALL IJCEN(IMAG,N,IC,JC)

IC,JC : output center of image

3.2 Centered transformation of image

The call   CALL TCEN(IMAG,CIMAG,N,M,IC,JC)

IC,JC : input coordinates of center

CIMAG : output image of size MxM

3.3 Determination of the rotation angle of image

The call   CALL TETA1(IMAG,N,T)

T : output rotation angle of image

3.4 Rotation of image

The call   CALL TROTA(IMAG,RIMAG,N,T,IT,JT)

RIMAG : output rotated image

T,IT,JT : given rotation angle and center

3.5 Scale of image

The call   CALL SCALEØ(IMAGØ,IMAG,SIMAG,N,IC,JC)

IMAG,IMAGØ : input images of size NxN centered at IC,JC

SIMAG : output scaling image

4. PROGRAMS FOR TRANSFORMATIONS OF IMAGE

4.1 Polcar Transform

The call   CALL TPOCAR(IMAG,IA,N)

IA : output image of size NxN modified by Polcar transform

4.2 Mellin transform

The call MELIN(IMAG,FA,N,M)

FA : output image of size MxM modified by Mellin transform

4.3 Fourier transform

The call CALL TFØ(IMAG,FA,N)

FA : output image of xize NxN amplitude spectrum image

5. PROGRAMS FOR FINE MATCHING

5.1 Computation of invariant moments of window

The call CALL MOMENT(IMAG,K,I,J,KSI)

K : size of window

KSI : output vecotr consisting of loga of 7 invariant moments

5.2 Computation of kL-centroid of window

The call CALL TOPO(IMAG,I,J,CENTER,GRLV)

GRLV : gray level of image

CENTER 2, GRLV : output array

CENTER(1,1), CENTER(2,1) is the global center of image

CENTER(1,K), CENTER(2,K) are centers corresponding to the region of gray level K, K=1,...GRLV-1.

A kép-összeillesztő algoritmusok hatékony szintézise.

HOANG KIEM, PHAM NGOC KHOI

Összefoglaló

A szerző a kép-összeillesztő algoritmusok szintézisének hatékonyságára hierarchikus sémákat és "durva-finom" összeillesztő algoritmusokat használ. Tárgyalja a kép egyes tartományaira vonatkozó jellemző vonások kivonata-módszert, valamint a szintér-összeillesztő módszereket.

Эффективный синтез алгоритмов прикладывания образов

Хоанг Кием, Пхам Кхои

Резюме

Для повышения эффективности синтеза алгоритмов прикладывания образов авторы используют иерархические схемы и алгоритмы типа "грубо-тонкие". В статье дискитируются также некоторые другие методы.