International Journal on Emerging Technologies **1**(2): 1-4(2010) **ISSN : 0975-8364**

On convergence of hopfield neural networks for real time image matching

Varun Kumar, Ashish Chaturvedi*, M.P. Singh** and M.K. Gupta***

Dept. of Applied Sciences, VCE, Meerut,
**Dept. of Applied Sciences GBIT Meerut,*
***Dept. of CS, ICIS, Dr. B. R. Ambedkar University, Agra (UP);*
****Dept. of Mathematics, C. C. S. University, Meerut, India*

ABSTRACT : Present paper demonstrates on innovative approach for a fundamental problem in computer vision to map real time a pixel in one image to a pixel on another image of the same scene, which is generally called image correspondence problem. It is a novel real time image matching method which combines Rotational Invariant Feature Selection for real time images and optimization capabilities of Hopfield Neural Networks. The most invariant image matching features are extracted from the reference image. Finally, the image matching process is optimized by Hopfield neural networks, where image matching problem is treated as minimization of energy function of the Hopfield neural networks.

Keywords : Image correspondence, Hopfield model, Rotational Invariant Feature Selection.

I. INTRODUCTION

The image matching problem, also known as the image correspondence problem [1], is one of the most exigent tasks in the computer vision research field. It becomes more challenging when we perform image matching under varying conditions for virtually intelligent vision systems including automated image registration, object recognition and tracking, content based database retrieval and image based modeling [2, 3, 4,]. Many problems such as camera calibration [5], [6], [7], 3-D object reconstruction [8], obstacle detection [9], [10], motion estimation [11] and object tracking [12], require solving the correspondence problem as an initial step for processing the sequence of further steps. The matching problem can be defined as the establishment of the correspondence between features extracted from two or more images of the same scene. However, the matching problem is a well known ill-posed problem. The solution of correspondence problem may not exist if a point in one image does not have a corresponding match in the other image due to occlusion. The solution of the problem may not be unique if there may be more than one match due to scenes with repetitive patterns.

Traditionally, the two types of techniques are used to solve the problem of image correspondence: area-based and feature-based techniques. Area-based techniques utilize correlation between the intensity patterns in the neighborhood of a pixel in the reference image and those in the neighborhood of a corresponding pixel at the real-time image. Feature-based techniques, on the other hand, using symbolic features derived from intensity images, such as edge points and edge segments, allow simple comparisons between the attributes of features. In this way, the feature based methods are generally faster than area-based methods.

Feature selection has been widely used to reduce computation time and improve accuracy. Multi-class SVM was used in [13] to select the most informative features for face recognition. The proposed SVM-DFS can speed up classication without degrading the matching accuracy. Mahamud and Hebert [14] proposed discriminative object parts selection and used conditional risks as the distance measure in nearest neighbor search. Dorko and Schimid [15] introduced a method for selecting most discriminative object-part classiers based on likelihood ratio and mutual information. None of these approaches focuses on rotation invariance or utilizes the additional information introduced by specically designed and labeled training views.

Many local feature based image matching systems utilize orientation alignment for rotation invariance. In orientation histograms are computed from local circular regions of the relative scale. However, histogram based methods are computationally too expensive to be used in real time image matching systems, because the process generally involves time-consuming steps such as relative scale searching, dominant orientation calculation, and pixel values extraction from irregular regions [3].

To attempt such problem, Lepetit and Fua [2] introduced a simplified orientation correction technique for real-time image matching applications. The method only considers intensity changes along a fixed-size circular region centered on each key point. It is not, however, robust to scale changes or out-of-plane rotations.

To enhance rotation invariance in real-time image matching systems, we are using an approach combining feature selection and multiple-view training into one unied framework. Firstly, we construct a small number of rotation-dominant views and obtain a set of descriptors for each view track. Then for those feature points with a high repeatability, Raw Ranking Scores (RRS) are calculated based on feature distinctiveness and invariance. Finally, the raw ranking score is rescaled, weighted and combined with the other traditional feature selection criterion for the Final Ranking Score (FRS). Features with high FRS are selected.

None of the feature extraction algorithms available today is capable of avoiding data loosing, especially under the influence of distortion factors. Thus the result of image matching is affected greatly [16]. To conquer the disturbances mentioned above, this paper demonstrate a novel image matching method, which combines Rotational Invariant Feature Selection and the optimization capability of Hopfield neural networks [17].

II. A REAL TIME FEATURE EXTRACTION METHOD: ROTATIONAL INVARIANT FEATURE SELECTION

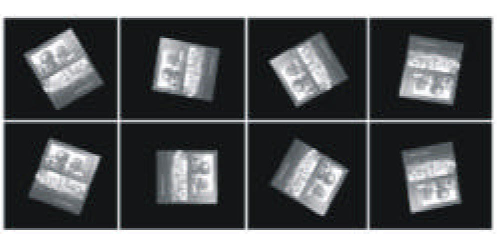
We can extract invariant features under various rotations by generating small number of rotational domains synthesized views with the help of affine transformation [18] for the purpose of pre-processing.

$$I' = U_A I = \begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix} I$$

$$A = B(\Omega)\, R(-\phi) D R(\phi) \text{ and } D = \begin{bmatrix} \xi_1 & 0 \\ 0 & \xi_2 \end{bmatrix}$$

The rotation parameter Ω uniformly distributes from – p to p. Other parameters ranges are: $\phi \in [-\pi/2, \pi/2]$, $\xi_1, \xi_2 \in [0.95, 1.05]$, t_1, t_2 = 0 or 1.

The rotational dominant synthesized views cover all the possible rotations; microscopic changes as well as macroscopic changes. We can also generate scale-domain synthesized views for scale invariant feature selection. This feature selection method can be depicted from the flow diagram shown below.



Rotational View Domains

The feature selection method can be described as:

Step 1: Construction of set of local patches: Firstly, let us assume that the numbers of training views N_v and feature points N_F are separately observed in these views. Also V_P ($p = 0 - N_V$), F_q ($q = 1 - N_F$) represent training view and descriptor N_D for those feature points respectively. Because the affine transformations from V_0 (the input image) to all the synthesized training views are known, we are able to identify subsets of F_q belonging to the same physical locations. Such a subset is called a view track of the object, represented by S_k ($k = 1 - N_s$). In general we can write:

$$\delta_{i,k} = \begin{cases} 1 & if\ featurebelongstoviewtrackk \\ 0 & otherwise \end{cases}$$

In this way we can describe 32 × 32 local patches around feature points. Our approach is to treat each patch as a vector composed of 1024 pixel intensities. Using the Walsh-Hadamard kernels, these vectors will be projected onto a space with much lower dimensionality (N_D). These W.H. kernels projection approach are reliable even under very noisy conditions and fast enough for real-time systems. The compact feature representations and $\delta_{i,k}$ values for all the feature points and view tracks are the outputs of view set construction.

Step 2: Selection of invariant features: After constructing the view tracks, our next step is to select view tracks that are unique, invariant and constant. The stability is measured by the feature repeat rate across all the training views, equivalent to the size view tracks. The stability score for view track k is defined as : $SS_k = \sum_{1 \leq i \leq N_F} \delta_{i,k}$. Feature points having stability scores lower than the threshold value (L_{SS}) are eliminated. Features that are not distinguishing from one view track to another are usually come from repeated elements of the scene, cause confusion to image matching system. While training performed for rotational view domains, features that have a high possibility to come from boundaries of a region, forming more diverse descriptors under rotation and potentially compromising the system's robustness. Hence, we have to select those feature points that are unique and invariant, and accordingly to improve the geometric invariance of the system.

Step 3: Computation of ranking scores for each view track: The expected value of distance between view track mean vectors will compute the uniqueness of the features. In this way, we will measure the variance of each view track by the expected value of single dimension variances. Thus, for ranking the features, Mean and variance must be calculated.

To measure the uniqueness of one view track U_j, we first compute the mean vector of all the feature vectors belonging to U_j'

i.e. $$M_j = \frac{\sum_{i=1}^{N_F} \delta_{i,j} F_i}{\sum_{1=i}^{N_F} \delta_{i,j}}$$

Let $D_{j:j}'$ represent the mean vector distance between view track U_j and D_j'. The Uniqueness Score for view track U_j can be demonstrated as:

$$US_j = \frac{1}{N_U} \sum 1 \le j \le N_U D_{j:j'}$$

However, computing co-variance matrices for each view track is time-consuming, we are here simplifying variance computation by individually observing each dimension of all feature vectors of one view track, which provides a N_D–dimensional variance vector for each view track i.e.

$$VV_{j,l} = \frac{\sum_{i=1}^{N_F} \delta_{i,j} (F_{i,l} - M_{j,l})^2}{\sum_{i=1}^{N_F} \delta_{i,j}}$$

where $l = 1 - N_D$ is the dimension of vectors.

The variance Score for view track U_j is defined as the expected value of all VV_j 's components:

$$VS_j = \frac{1}{N_D} \sum_{1 \le l \le N_D} VV_{j,l}$$

Good features should have high uniqueness value and low variance. Therefore, the Raw Rank Score (RRS) for view track U_j is defined as: $RRS_j = \dfrac{DS_j}{VS_j}$.

The unified formula for RRS expressed by original feature descriptors is:

$$RRS_j = \frac{\frac{1}{N_T} \sum_{j'=1}^{N_T} ! \sum_{i=1}^{N_F} \delta_{i,j} F_i - \sum_{i=1}^{N_F} \delta_{i,j'} F_i !}{\frac{1}{N_D} \sum_{l=1}^{N_D} \frac{\sum_{i=1}^{N_F} \delta_{i,j} (F_{i,l} - \sum_{i=1}^{N_F} \delta_{i,j} F_{i,l})^2}{\sum_{i=1}^{N_F} \delta_{i,j}}}$$

Finally, RRS is rescaled to the same range as stability score and combined together through a weight parameter ($\alpha = 0 - 1$) to form Final Ranking Score.

$FRS_j = \alpha RRS_{rescalled}, j + (1 - \alpha) SS_j$. Here a = 0 is the traditional criterion when only repeatability is considered while a = 1 is the extreme case using only RRS.

Step 4: Selection of view track of high ranking score: Ranking all the view tracks, final step is to select the object features with scores higher than some threshold value.

III. OPTIMIZING IMAGE MATCHING PROCESS BY HOPFIELD NEURAL NETWORK

The key idea of image matching by Hopfield neural network is to seek an appropriate energy function expression for the problem, so as to make the Hopfield network convergence state corresponding with an image matching result.

The energy function of Hopfield model for some current state of the images can be given as

$$E = -\frac{1}{2} \sum_i \sum_j W_{ij} O_i O_j - \sum_i O_i \theta_i \qquad ...(1)$$

where, = Adjoining weights between i^{th} and j^{th} neuron

O_i & O_j = output states of i^{th} and j^{th} neuron

θ_1 = external inputting threshold acting on i^{th} neuron.

Now if suppose the reference image high RRS feature set size is P, the real time image High RRS feature set size is Q, obviously, the latter just corresponds to a subset of the former, so Q is much greater than P. Then the answer of Hopfield neural network image matching can be expressed by an $P \times Q$ neuron matrix $\{O\}$, where $O_{i,j}$ ranges in [0,1], and the neuron state O converges to 1 if the feature from view track i matches the image feature from view track j perfectly; otherwise it is equal to or close to 0. To the matrix $\{O\}$ mentioned above, due to the uniqueness of image, number of "1" in each row and each column of the matrix should be not more than one.

The Hopfield energy function for matching an image with its reference image can be derived as

$$E = \frac{\alpha_1}{2} \sum_x \sum_i \sum_{j \ne i} W_{xi} W_{xj} + \frac{\alpha_2}{2} \sum_i \sum_x \sum_{y \ne x} W_{xi} W_{yj} + \frac{\alpha_3}{2} \left[\sum_x \sum_i W_{xi} - \min(m,n) \right]^2 + \frac{\alpha_4}{2} \sum_x \sum_i \{\text{Value}(x) - \text{Value}(i)\}^2 W_{xi} + \frac{\alpha_5}{2} \sum_x \sum_i [(US_i)_1 - (US_i)_2]^2 W_{xi} \qquad ...(2)$$

Where, α_1, α_2, α_3, α_4 and α_5 are the weight coefficients. The first term of the equation is compatibility constraints, second term is uniqueness constraints, third, and fourth and fifth terms are matching constraints. It has already been proved that the Hopfield neural network with such an energy function has a poor convergence. Among 100 times' experiments done by Wilson, only 15 times' experimental results are convergence to mean answer within 1000 times' iterative computation [17].

As seen from the energy equation, the matching process tries to satisfy the constraints globally in a parallel manner to find the matching nodes. The local compatibility measures provide excitory and inhibitory supports for matching local features which are unique for the images. The dynamics of the network is characterized as a stochastic process that will reach a stable state when the energy Eq. (4) is at its minimum.

IV. CONCLUSION

The paper demonstrates a robust method for real time image registration using Hopfield Neural Networks. Rotational invariant features can be scaled and selected from the real time reference image. This reference image can be matched with a real time image by converging Hopfield energy equation.

REFERENCES

[1] Weng, J., Ahuja, N. & Huang, T. Two-view matching. *In Proceedings of Int. Conf. on Computer Vision,* 65–73, (1988).

[2] Lepetit, V. & Fua, P. Keypoint recognition using randomized trees. Transactions on Pattern Analysis and Machine Intelligence. **28**(9)**:** 1465-1479 (2006).

[3] Lowe, D. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision.* 91–110 (2004).

[4] Wang, Q., Mooser, J., Neumann, U. & You, S. 2008. Augmented exhibitions using natural features. *International Journal of Virtual Reality.* **18**(2)**:** 46-65 (2008).

[5] Vigueras, J.F., Berger, M.O. & Simon, G. Iterative multi-planar camera calibration: improving stability using models election. *International Journal of Vision,* Video and Graphics. 1-8 (2003).

[6] Matsunga, C. & Kanatani, K. Calibration of a moving camera using a planar pattern: optimal computation, reliability evaluation and stabilization by model selection *In Proceedings of 6th Euro. Conf. Computer Vision, Dublin, Ireland* (2)**:** 595–609 (2000).

[7] Zhang, Z. Flexible camera calibration by viewing a plane from unknown orientations. In. In Proceedings of 7th Inter national Conference on Computer Vision. **1:** 666-673 (1999).

[8] Longuet, H.C. A computer algorithm for reconstruc-ting a scene from projections. *Nature,* **293**: 133-135 (1981).

[9] Zhang, Z., Weiss, R. & Hanson, A.R. Obstacle detection based on qualitative and quantitative 3d reconstruction. IEEE trans. on Pattern Analysis and Machine Intelligence. **19**(1)**:** 15-26 (1977).

[10] Elias, R., Haijdiab, H. & Lagani, R. Wide baseline obstacle detection and localization. In Proceedings of the *IEEE Seventh Inter national Symposium on Signal Processing and its Applications, Paris-France,* **1:** 21-24 (2003).

[11] Fornland, P. Direct obstacle detection and motion from spatio-temporal derivatives. *In CA IP , Prague, Czech Republic* (1995).

[12] William, T. & Thorpe, C. A specialized multi baseline stereo technique for obstacle detection. *In IEEE conf. on Computer Vision and Pattern Recognition,* 238-244 (1998).

[13] Fan, Z. & Lu, B. Fast recognition of multi-view faces with feature selection. In IEEE International Conference on Computer Vision. 76-81 (2005).

[14] Mahamud, S. & Hebert M. The optimal distance measure for object detection. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 245-255 (2003).

[15] Dorko, G., & Schmid, C. Selection of scale-invariant parts for object class recognition. *In IEEE International Conference on Computer Vision.* 634-639 (2003).

[16] Zhenghao, S. Huang, S. Artificial Neural Network Image Matching, Microelectronics and Computer, **20**(Supl.)**:** 18-21 (2003).

[17] Hopfield J. & Tank, D.W. Neural Computation of Detection in Optimization Problems. *Biological Cybernetics.* (36)**:** 141-152 (1985).