New Feature Correspondence Method using Bayesian Graph Matching Algorithm

Wan Hyun Cho¹, In Seop Na²*, Sun Worl Kim¹, Soon Ja Kang³
¹Department of Statistics, Chonnam National University, Gwangju, 500-757, Korea
²School of Electronic & Computer Engineering, Chonnam National University, Gwangju, 500-757, Korea
³Department of Mathematics Education, Chonnam National University, Gwangju, 500-757, Korea

Abstract - In this paper, we propose a new approach for establishing the correspondence between local invariant features using Bayesian graph matching algorithm. First, we will discuss various local invariant feature detectors and descriptors for scale and affine transformation and illumination changes. Second, we propose an efficient features corresponding method using local invariant features and new graph matching algorithm. Finally, we evaluate the comparative performances of our method (BGMA) with several existing feature matching approach such as Lowe’s shift matching by conducting experiments on various real image data. We test on three image pairs taken from MSRC v2 dataset and Caltech 101 dataset. Experimental results show that the proposed method clearly outperforms rather than the existing matching algorithms about feature correspondence in images with rotation or scale transformation and illumination changes.

Keywords: Bayesian, Graph Matching, Correspondence, Invariant Feature

1. Introduction

Image feature detection and matching algorithm is a fundamental task in image processing and machine vision. A various local features representation allows us to exploit model relationships between different objects, or to recognize the disparity between difference objects under consideration. The local features on distinguished region of given image are usually represented by detectors and descriptors. Normally key-points or detectors represent positions or anchor locations of image feature, object parts or elementary units of consideration for the task at hand, and descriptors describe a structure relationship of detectors or the local appearance in their neighborhoods. Here, feature matching is the task of establishing a local features correspondence such as the detectors and descriptors between two given images such that some constraints are satisfied. But, this problem has been referred to as the most difficult part of 3D structure recovery, and is particularly challenging if the images have been taken form widely separated viewpoints.

Many features matching algorithms have been proposed during last few decades [1-3]. Among these methods, the similarity measure is one of the most powerful tools for feature matching. In order to find the corresponding point for a feature point using the similarity measure, a template window is considered around the feature point and this window is shifted pixel by pixel across a larger search window around an estimated corresponding point, and in each position the similarity between the two regions is measured. The maximum or minimum value of the resultant measurements defines the position of the best match. Normalized cross correlation and SSD (sum of squared differences) are well-known methods for measuring similarity between two regions. In addition to a normalized similarity value, normalized cross correlation has the advantage of being invariant to the linear change between the data sets, which makes the algorithm robust against low varying illumination change the scene. In recently, Torresani et al.[4] presented a new approach for establishing correspondences between spare image features. They formulate this matching task as an energy minimization problem by defining a complex objective function of appearance and the spatial arrangement of the features and used a novel graph matching optimization technique, which they refer to as dual decomposition.

In this paper, we proposed new approach for matching two sets of local invariant features using Bayesian graph matching algorithm. First, we will discuss various local invariant features such as detectors and descriptors for scale and affine transformation and illumination changes. Second, we propose the Bayesian graph matching algorithm with local and global properties and then we induce method that can be efficiently matching for local invariant features using our algorithm. Finally, we have conducted the comparative experiments on various real images.

2. Local Invariant Features

A local feature on distinguished region of given image represents an image pattern which differs from its immediate neighborhood. It is usually associated with a change of an image property or several properties simultaneously. The image properties commonly considered are intensity, color and texture. Local features can be points, edges or small image patches. They are usually represented by detectors and descriptors. The detectors represent normally positions or anchor locations of image feature, object parts or elementary units of consideration for the task at hand, and then the descriptors describe a structure relationship of detectors or the local appearance in their neighborhoods. Here, we have discussed various detectors and descriptors describing the image characters that are locally
invariant with image rotation, scale transformation and illumination changes.

2.1 Local invariant feature detector

As most of the object recognition system based on the local appearance, it is of great importance to find distinguished regions in a highly repetitive manner. A distinguished region is a connected part of an image showing a significant and interesting image property. It is usually determined by the application of a region of interest detector to the image. If a region detector returns only an exact position within the image we also refer to it as interest point detector. The most important information that ideal region detectors give to us is the location of features, but other characteristics such as shape (scale) and orientation of a region of interest have to deliver additionally. Moreover, two important characteristics that need to be a good detector are repeatability and reliability. Repeatability means that the same feature can be detected in different images. Reliability means that the detected point should be distinctive enough so that the number of its matching candidates is small.

The currently most popular distinguished detectors can be roughly divided into three categories; corner based or intensity based, contour based or region based, parametric model based detector or other approaches [5,6,7]. Corner based detectors locate interest points and region which contain a lot of image structure such as edges or intensity changes around points. To measure the change, first and second derivatives of images are used in many different forms and combination. Therefore, these are not suited for uniform regions and region with smooth transitions. Region based detectors regard local blobs of uniform brightness as the most salient aspects of an image and extract contours from image. Parametric model based detectors or other approach take into account the entropy of a region or try to find interest points by matching models or templates to an image. The most popular interest detectors, which give sufficient performance results, are listed: Harris or Hessian point based detectors (Harris, Hessian, Harris-Laplace, Hessian-Laplace), Difference of Gaussian Points (DoG) detector, Entropy Based Salient Region (EBSR) detector, Harris or Hessian affine invariant region detectors (Harris-Affine, Hessian-Affine), Maximally Stable Extremal Regions (MSER) detector, and Edge Based Regions (EBR) detector and Intensity Based Regions (IBR) detector.

2.2 Local invariant feature descriptor

Next, we have also discussed the interest feature descriptors and their characteristics. A descriptor is a process that takes information of features and image to produce descriptive information i.e. features’ description, which are usually presented in form of features vectors. The descriptions then are used to match a feature to one in another image. And two important aspects that a descriptor has to satisfy are discriminative and invariant. First, the descriptors should be discriminative enough to distinguish between features of the images stored in the database. That is, handling thousands of database images requires the ability to distinguish between a vast numbers of descriptors, demanding thus highly discriminative representation. Second, a descriptor has to be invariant or at least in some degree robust to variations in an different view images that are not reflected by the detector. It is invariant to rotation, scaling, and affine transformation so the same feature on transformed images will be characterized by almost the same value and distinctive to reduce number of possible matches.

The currently most popular distinguished descriptors can be roughly divided into three categories; distribution based descriptors, filter based descriptors, and other methods [8,9]. Distribution based descriptors use histograms to represent the characteristics of the region. The characteristics could be pixel intensity, distance from the center point, relative ordering of intensity, or gradient. Filter based descriptors use a set of differential operators or different types of filters to describe an interest region. Other approach takes into account the moments or gradient moment with high order degrees to characterize shape and intensity distribution in an interest region. Many different techniques for describing local image regions have been developed. They are listed: Scale Invariant Feature Transform (SIFT) descriptor, PCA-SIFT, Gradient location and orient histogram (GLOH), shape context, spin images, differential invariants, steerable filters and complex filters, and moment invariants.

3. Efficient Feature Matching Method

3.1 Formulation of Probabilistic Feature Matching Algorithm

Here, we will formulate feature matching problem as Bayesian inference framework. The feature matching implies the correspondence between two sets of feature detectors \( P(1), P(2) \) using their descriptor properties \( D(1), D(2) \) given from the two input \( I(1) \) and \( I(2) \). This is represented by the bi-directional detector to detector correspondences mapping and \( m : P(2) \rightarrow P(1) \).

And also it is efficiently represented \( m : p(1) \rightarrow p(2) \) using an assignment matrix \( X \in \mathbb{R}^{N_1 \times N_2} \) of nonnegative real numbers, where \( N_1 \) and \( N_2 \) denote the numbers of feature detectors in \( P(1) \) and \( P(2) \) respectively. That is, each component \( X_{(i,j)} \) of the assignment matrix \( X \) is representing the possibility that a detector \( p(i) \in P(1) \) matches to some detector \( p(j) \in P(2) \). Hence, in order to find the optimal assignment matrix, we are going to apply the Bayesian inference principle with the
matching problem between two sets of feature detectors $(P(1), P(2))$.

3.1.1 Prior distribution of assignment matrix

First, we will consider the prior distribution for row and column vectors of the assignment matrix $X$. Here, we factorize an assignment matrix $X$ as the $N_1$ row vectors $x_1^R, \ldots, x_{N_1}^R$ and the $N_2$ column vectors $x_1^C, \ldots, x_{N_2}^C$. Then, each component $x_{ij}$ for one of row vector $(x_i^R)^T = (x_{i1}, \ldots, x_{ij}, \ldots, x_{iN_2})$ represents the possibility that the detector $p_i^{(1)} \in P(1)$ is mapping to any one of detectors $p_j^{(2)} \in P(2)$. Hence, we can assume that these components satisfy two properties. These are $x_{ij} \geq 0$ for all $j = 1, \ldots, N_2$ and $\sum_{j=1}^{N_2} x_{ij} = 1$. Therefore, we can define a new discrete random variable $Y$ taking with finite number of values $1, 2, \ldots, N_2$. And the probability distributions of $Y$ taking $y$ can be parameterized by the component of a vector $x_i^R$, that is, $p(Y = y) = x_{ij}, y = 1, \ldots, N_2$. Another way to write this is

$$p(y | x_i) = \prod_{j=1}^{N_2} x_{ij}^{(y=j)}$$

where $I(y = j)$ denote an indicator function. Moreover, we assume that a conjugate prior for each row probabilistic vector $x_i^R$ is the Dirichlet distribution:

$$p(x_i^R | \alpha) \sim D(\alpha_1, \ldots, \alpha_{N_2}) = \frac{\Gamma(\sum_{j=1}^{N_2} \alpha_j)}{\prod_{j=1}^{N_2} \Gamma(\alpha_j)} \prod_{j=1}^{N_2} x_{ij}^{\alpha_j-1}$$

where $\Gamma(\cdot)$ denotes the Gamma function and $\alpha_1, \ldots, \alpha_{N_2}$ are positive numbers. The hyper-parameter $\alpha_j$ can be interpreted as a virtual occurrence for value $x_{ij}$. Large $\alpha_j$ corresponds to strong prior knowledge about the distribution and small $\alpha_j$ corresponds to ignorance. Then, using properties of the Gamma distribution and Dirichlet distribution, we can generate the pseudo random variable for the $(i, j)^{\text{th}}$ component $x_{ij}^R$ of each row probabilistic vector $x_i^R$ as follows:

$$x_{ij}^R \mid (\alpha_1, \ldots, \alpha_{N_2}) \sim \frac{\Gamma(\alpha_j, \beta)}{\sum_{k=1}^{N_2} \Gamma(\alpha_k, \beta)}$$

3.1.2 Likelihood function for assignment matrix

Second, in order to solve the feature matching in a Bayesian inference principle, we need to construct the $(N_1 \times N_2)$ likelihood function matrix $L$ with likelihood components $l_{i,j}(p)$ that represents the possibility of some detector $p_i^{(1)} \in P(1)$ matching to any detector $p_m^{(2)} \in P(2)$ for the assignment matrix $X$. It can be derived from the weighted combination of the matching matrix $Y$ that is induced from the similarity for a pair of descriptors defined at two detectors and the distance matrix $D$ that represents the spatial distance between locations of detectors. First, we consider the matching matrix $Y$ that is induced from the similarity for a pair of descriptors corresponding two detectors. We define edges $e_{ij}^{(1)} \in E^{(1)}$ and $e_{ab}^{(2)} \in E^{(2)}$ as a pair of detectors $(p_i^{(1)}, p_m^{(2)})$ and $(p_j^{(2)}, p_k^{(2)})$. For each pair of $e_{ij}^{(1)} \in E^{(1)}$ and $e_{ab}^{(2)} \in E^{(2)}$, the
similarity matrix $S$ of size $N_1 \times N_1 \times N_2 \times N_2$ is defined by having its elements as similarities:

$$S(e_{ij}^{(1)}, e_{ab}^{(2)}) = f((d_1^{(1)}, d_2^{(1)}), (d_2^{(2)}, d_1^{(2)}))$$  \hspace{1cm} (6)

Here the similarity function $f(\cdot)$ can be labeled with an inverse function of the Euclidean distance between their two edges. That is $f((d_1^{(1)}, d_2^{(1)}), (d_2^{(2)}, d_1^{(2)})) = \exp(-||d_1^{(1)} - d_2^{(1)}|| - ||d_2^{(2)} - d_1^{(2)}||)$.

Then, we can compute the $(N_1 \times N_2)$ matching matrix $Y$ with components $y_{(i,m)}$ that represents the possibility of some detector $p_i^{(1)} \in P^{(1)}$ matching to any detector $p_m^{(2)} \in P^{(2)}$ from the similarity matrix $S$ as following form. That is, it can be obtained by summing the similarities for all pairs of edges containing detectors $p_i^{(1)} \in P^{(1)}$ and $p_m^{(2)} \in P^{(2)}$:

$$Y = (y_{(i,m)}), \quad y_{(i,m)} = \sum S(e_{ij}^{(1)}, e_{ab}^{(2)})$$  \hspace{1cm} (7)

Second, we think about the distance matrix $D$ that represents the spatial distance between positions for a pair of detectors contained two images. In matching problem for two detectors, we have to consider the local information which contains detectors $p_i^{(1)} \in P^{(1)}$ and $p_m^{(2)} \in P^{(2)}$.

$$\text{D} = (d_{(i,m)}), \quad d_{(i,m)} = \exp(-\frac{||p_i^{(1)} - p_m^{(2)}||}{\max_{p_i^{(1)}, p_m^{(2)}} ||p_i^{(1)} - p_m^{(2)}||})$$  \hspace{1cm} (8)

Here, the large value of $d_{(i,m)}$ amplifies the matching potentiality but the small value of $d_{(i,m)}$ attenuates this property.

Finally, we combine the matching matrix and the distance matrix to construct the likelihood function matrix $L$. First, in order to agree with scales for elements of the matching matrix $Y$ and the distance matrix $D$, we apply the bi-stochastic normalization scheme with two matrices. Next, adopting the proper weighting parameter $\alpha$, the likelihood function matrix $L$ is obtained by using the following equation:

$$L = \pi Y + (1 - \pi)D, \quad 0 \leq \pi \leq 1$$  \hspace{1cm} (9)

Hence, we use the $i^{th}$ row and $j^{th}$ column element $l_{i,j}$ of the likelihood matrix $L$ as the likelihood function that represents the possibility of some detector $p_i^{(1)} \in I^{(1)}$ matching to any detector $p_j^{(2)} \in I^{(2)}$. That is, we have that

$$l_{i,j} = \text{Likelihood}(p_i^{(1)} \text{ matching } p_j^{(2)}), i = 1, \cdots, N_1, j = 1, \cdots, N_2$$  \hspace{1cm} (10)

3.1.3 Posterior distribution of assignment matrix

Finally, by combining the prior distribution $p(x_i^{(1)} | \alpha)$ and the likelihood function $L((p_i^{(1)}, p_j^{(2)}) | x_{ij})$ of $(i, j)$ component $x_{ij}$ of row vector $x_i^R$ and column vector $x_j^C$ using Bayes formula, we have obtained the posterior distribution $p(x_i^{(1)} | x_j^{(1)}, L_{ij})$ of $(i, j)$ component $x_{ij}$ of row vector $x_i^R$ and the posterior distribution $p(x_j^{(2)} | x_i^{(2)}, L_{ij})$ of $(i, j)$ component $x_{ij}$ of column vector $x_j^C$ respectively as follows:

$$p(x_i^{(1)} | x_j^{(1)}, L_{ij}) = \frac{p(x_i^{(1)} | \alpha, L_{ij}) \cdot L((p_i^{(1)}, p_j^{(2)}) | x_{ij})}{\sum_{k=1}^{N_2} p(x_i^{(1)} | \alpha, L_{ik}) \cdot L((p_i^{(1)}, p_k^{(2)}) | x_{ik})}$$  \hspace{1cm} (11)

and

$$p(x_j^{(2)} | x_i^{(2)}, L_{ij}) = \frac{p(x_j^{(2)} | \alpha, L_{ij}) \cdot L((p_j^{(2)}, p_i^{(1)}) | x_{ji})}{\sum_{k=1}^{N_1} p(x_j^{(2)} | \alpha, L_{kj}) \cdot L((p_j^{(2)}, p_k^{(1)}) | x_{kj})}$$  \hspace{1cm} (12)

Here, we have repeated this algorithm iteratively by taking the posterior probabilities $(x_i^{(1)} | \alpha, L_{ij})$ and $(x_j^{(2)} | \alpha, L_{ij})$ as initial value of the hyper-parameter $\alpha$ again. If the difference between values of one step before and after step can be ignored, we will stop the iteration. Finally, we have obtained that the row and column matching matrices $X_R$ and $X_C$ between two features $(p_i^{(1)}, p_j^{(2)})$ are defined by

$$X_R = \left[ \begin{array}{cccc} x_1^R & \cdots & x_{N_1}^R \\ \vdots & \ddots & \vdots \\ x_{N_1}^R & \cdots & x_{N_1}^R \end{array} \right] \quad \text{and} \quad X_C = \left[ \begin{array}{cccc} x_1^C & \cdots & x_{N_1}^C \\ \vdots & \ddots & \vdots \\ x_{N_1}^C & \cdots & x_{N_1}^C \end{array} \right].$$  \hspace{1cm} (13)
3.1.4 Computation of optimal feature matching solution by convex problem relaxation

Next, we consider an optimization problem for recovering the optimal solution $X^*$ from probabilistic matching matrices $R$ and $C$ by minimizing the distance between $X$ and $R$ or between $X$ and $C$ defined by the following form:

$$X^* = \arg\min_{X \geq 0} \left( D_{KL}(X \| R) \right) \quad \text{or} \quad X^* = \arg\min_{X \geq 0} \left( D_{KL}(X \| C) \right) \quad \text{s.t.} \quad X \leq 1, \quad X^T 1 \leq 1, \quad 1^T X 1 = k. \quad (14)$$

In this case, we are going to use the distance function as the Kullback-Leibler divergence that is a non-symmetric measure of the difference between two probability distribution $P$ and $Q$. It is defined to be

$$D_{KL}(P \| Q) = \sum_{ij} P(i, j) \log \frac{P(i, j)}{Q(i, j)} \quad (15)$$

Moreover, we assume that $k$ detectors are matching at two images.

This is a special case of the general convex semi-definite program (SDP) which has the following standard form:

$$\min \quad \text{Tr}[\hat{Q} X] \quad \text{s.t.} \quad \text{Tr}[A, X] = C_i \quad (16)$$

where $A$ and $X$ are $m \times m$ matrices with $m$ the total number of matching vertices $\min(\|V \|, \|V^2 \|)$ and the maximized value of the likelihood function is defined as the

$$P_f(H) = \arg\min_{X \in \mathcal{C}_i} f(X) - \langle X, H \rangle \quad (19)$$

where $\langle X, H \rangle = \sum_{(i,j)} A(i,j)B(i,j)$ is the dot product between two matrices. Then, by repeated applications of $P_f(H)$ in a cycle manner, we obtain a primal-dual block update algorithm which is defined as follows.

Initial Step: Define $X_j^{(0)}, Y_j^{(0)}, j = 1,2,3$ and set $X_j^{(0)} = 0$ and $Y_j^{(0)} = X^R, j = 1,2,3$.

Use the convention $X_j^{(0)} = X_j^{(r-1)}$.

Iteration Step: Iterate on $t = 1,2,\ldots$ until convergence:

For $j = 1,2,3$:

$$X_j^{(t)} = P_f(H_j^{(t-1)} + \nabla f(X_j^{(t-1)})) \quad (20)$$

Finally, at convergence with $T$ iterations, the optimal solution is given by

$$X^* = \frac{1}{3} (X_1^{(T)} + X_2^{(T)} + X_3^{(T)}) \quad (18)$$

The algorithm employs successive Bregman projections and is derived using the framework of Frenchel Duality.

3.2 Determination of optimal number of matching features

Finally, we consider the searching problem for the optimal number of matching points between two sets of feature points. For some fixed value $k$, we define the score function as the Bayesian information criterion (BIC) used as a criterion for optimal model selection among a class of parametric models. The formula for the BIC is given as

$$BIC_k = -2 \ln L + k \ln n \quad (20)$$

where $n$ is the number of data points used in model selection and $L$ is the maximized value of the likelihood function for the estimated model. Here, we adopt the number of data points $n$ as the total number of matching vertices $\min(\|V \|, \|V^2 \|)$ and the maximized value of the likelihood function is defined as the
sum of similarities between hyper-edge containing the matching vertices. It is given as

\[ L = \sum_{(e^{(1)}, e^{(2)}) \in \mathcal{E}^{(1)} \times \mathcal{E}^{(2)}} H(e^{(1)}, e^{(2)}) \]  

(20)

Then, we search for the number of matching points to minimizing the value of BIC over \( 0 \leq k \leq \min(|V^{(1)}|, |V^{(2)}|) \).

4. Experimental results

In this section, we evaluate the comparative performances of our method (BGMA) with several existing feature matching approach such as Lowe’s shift matching (SM) [11] by conducting experiments on various real image data.

![Figure 1. Experiments on images from Tiger with scale transformation.](image1)

(a) BGMA: 15 correct matches out of 15

(b) SM: 15 correct matches out of 15

![Figure 2. Experiments on images from Office with rotation transformation.](image2)

(a) BGMA: 30 correct matches out of 30

(b) SM: 27 correct matches out of 30

(a)MGMA: 18 correct matches out of 18
Figure 3. Experiments on images from House with rotation and translation transformation.

We test on three image pairs taken from MSRC v2 dataset and Caltech 101 dataset. The candidate local features are generated using the SIFT detector and descriptor. Figure 1 shows that two methods can be perfectly matching for scale transformed image. But, as showed in Figure 2 and 3, our method is better matching than SM method in both the rotated image and the rotated and translated image.

5. Conclusions

We have presented the mathematical formula for the feature matching problem using both local invariant features and Bayesian graph matching algorithm. Our method considers the properties of detectors and descriptors extracted from given image. And it also induces the bidirectional detector to detector probabilistic matching algorithm. Main advantage of our algorithm is that it can be achieved by Bayes theorem and an iterative convex successive projection algorithm in order to the global optimum solution for feature matching problem. Experimental evaluations demonstrate that our method clearly outperforms the previous matching approaches on various real images.

Acknowledgements

This work was supported in part by the Korea Research Foundation Grant by the Korean government (KRF-2010-0013757).

References


