MVA'94 IAPR Workshop on Machine Vision Applications Dec. 13-15, 1994, Kawasaki

Region Correspondence for Color Scene Images Taken from Different Viewpoints

Caihua Wang
Graduate School of Electronic Science
Shizuoka University, Japan

Keiichi Abe
Department of Computer Science
Shizuoka University, Japan

3-5-1, Johoku, Hamamatsu
432 Japan

Abstract

A region-based matching approach is proposed for image correspondence of two images of the same scene but taken from different viewpoints. This approach consists of a matching stage and a validating and correcting stage. In the first stage, a local ternary relation among regions is used together with chromatic features of regions to establish region correspondences. In the second stage, the optical flow generated by the camera movement which can be estimated using all obtained correspondences is used to validate individual correspondences, and the ones which do not obey the optical flow are detected as mistakes and then corrected using the estimated optical flow.

Introduction

Image correspondence, that is, finding correspondences of various primitives between images of the same scene but taken from different viewpoints, is a basic task in many computer vision applications. The edge-based approach has been most widely used for image correspondence, but the region-based approach seems more promising. Using regions as matching primitives has the advantages that there are much fewer regions in an image than other primitives such as edges, that regions possess more information which supplies higher discriminating capability of regions, and that they are also more stable than, for instance, edges[1]. These advantages enable a more efficient and reliable matching by using regions, and the result can be used as guiding information for matching of more detailed primitives, such as edges, corners, etc.

There are a few methods proposed for region matching[1,2]. While the early one only used the features of regions for matching[1], the recent one tried to use structural information of region adjacency relations[2], which unfortunately tends to be meaningful and stable only for regions belong to the same object. In this paper we propose a region matching approach which consists of two stages: the matching stage and the validating and correcting stage. At the matching stage, local ternary relations, that is, the three triangles composed of the centers of gravity of the region in question and its three nearest neighbors, are introduced and all the regions are matched according to their chromatic similarity and local relational similarity. At the validating and correcting stage, the global consistency of the whole of obtained correspondences is utilized to detect and correct some apparent mistakes in the derived local matching results.

The matching stage

After the images were segmented into regions which consist of connected pixels with similar color, several features of each region, such as the average color, area, center of gravity, are computed beforehand. Then the following steps are carried out to establish region correspondences.

1. For each region A_i in one image, the corresponding candidate regions $\{B_{i_k}\}$ in the other image are selected as the ones which have color and area similar to A_i, that is, $\{B_{i_k}\}$ satisfy the following conditions

1) $1/T_a \leq a(A_i)/a(B_{i_k}) \leq T_a$ and

2) $D_c(A_i, B_{i_k}) \leq T_c$.

where $a()$ stands for area, and $D_c()$ is the *chromatic distance* in the modified HSV space[3] between two regions, defined as:

$$D_c(x, y) \equiv \sqrt{(V_x - V_y)^2 + S_x^2 + S_y^2 - 2S_xS_y\cos(H_x - H_y)}$$

T_c and T_a are certain thresholds and take the values of 0.07 and 4, respectively, in the experiments.

2. For each region A_i and its corresponding candidate B_{i_k}, compute their *chromatic similarity* by $S_c(A_i, B_{i_k}) = 1 - D_c(A_i, B_{i_k})/T_c$. The chromatic similarity is then normalized by

$$S(A_i, B_{i_k}) = S_c(A_i, B_{i_k})/\sum_j S_c(A_i, B_{i_j})$$

and is set to be the initial *matching similarity* between region A_i and B_{i_k}.

3. For each region A_i, local ternary relations formed by three triangles, which are composed of A_i and two of its three nearest neighbors A_i^1, A_i^2 and A_i^3, are constructed. In order to compute local relational

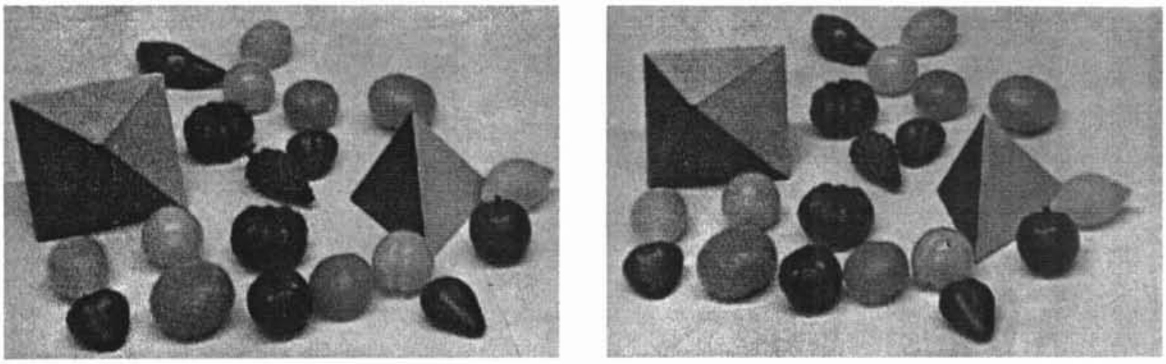

Fig. 1: Input images (only bright values are shown)

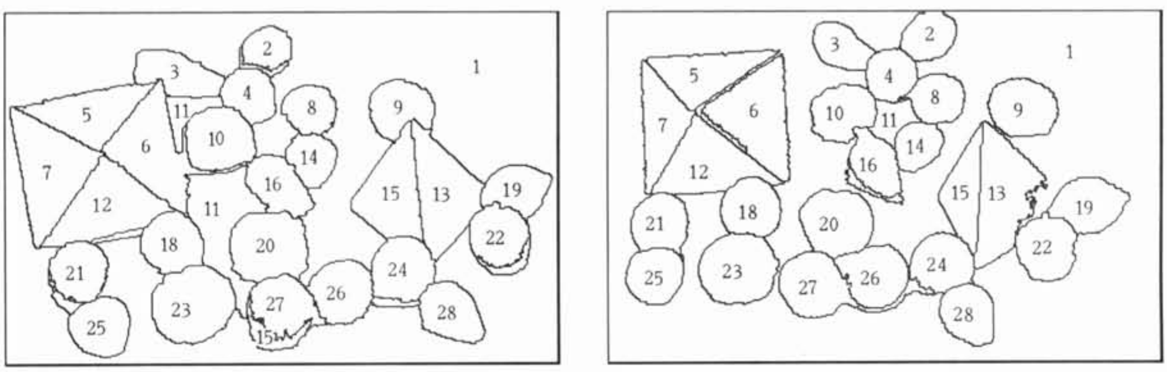

Fig. 2: Results of region segmentation and matching

similarity of region A_i and its corresponding candidate B_{i_k}, we map also A_i^1, A_i^2 and A_i^3 to their corresponding candidates $B_{i_k}^1$, $B_{i_k}^2$, $B_{i_k}^3$. This can be thought as mapping the three triangles to other three triangles in the other image. For each pair of a triangle and the mapped one, we compute their similarity in terms of edge similarity, angle similarity, and vertex (region) similarity. The sum of similarities of three pairs of the triangles is regarded as relational similarity between A_i and B_{i_k} under the specified mapping $\mathcal{P}$ of $(A_i^1, A_i^2, A_i^3) \xrightarrow{\mathcal{P}} (B_{i_k}^1, B_{i_k}^2, B_{i_k}^3)$. It is computed as:

$$\begin{aligned} S(A_i, B_{i_k}, \mathcal{P}) &= S_t(\triangle A_i A_i^1 A_i^2, \triangle B_{i_k} \mathcal{P}(A_i^1)\mathcal{P}(A_i^2)) + \\ &\quad S_t(\triangle A_i A_i^1 A_i^3, \triangle B_{i_k} \mathcal{P}(A_i^1)\mathcal{P}(A_i^3)) + \\ &\quad S_t(\triangle A_i A_i^2 A_i^3, \triangle B_{i_k} \mathcal{P}(A_i^2)\mathcal{P}(A_i^3)) \\ &= S_t(\triangle A_i A_i^1 A_i^2, \triangle B_{i_k} B_{i_k}^1 B_{i_k}^2) + \\ &\quad S_t(\triangle A_i A_i^1 A_i^3, \triangle B_{i_k} B_{i_k}^1 B_{i_k}^3) + \\ &\quad S_t(\triangle A_i A_i^2 A_i^3, \triangle B_{i_k} B_{i_k}^2 B_{i_k}^3) \end{aligned}$$

Here $S_t(\triangle abc, \triangle def)$ is the similarity between two triangles:

$$\begin{aligned} S_t(\triangle abc, \triangle def) = S_v(\triangle abc, \triangle def) S_e(\triangle abc, \triangle def) \\ S_a(\triangle abc, \triangle def) \end{aligned}$$

In this formula, $S_v(\triangle abc, \triangle def)$ is the chromatic similarity between two triangles, that is, the product of the chromatic similarities of three pairs of vertices of the two triangles:

$$S_v(\triangle abc, \triangle def) = S_c(a, d) S_c(b, e) S_c(c, f)$$

$S_e(\triangle abc, \triangle def)$ is the similarity of edges of the two triangles and is computed by

$$S_e(\triangle abc, \triangle def) = S_l(\overline{ab}, \overline{de}) S_l(\overline{ac}, \overline{df}) S_l(\overline{bc}, \overline{ef})$$

where S_l stands for similarity of length between two line segments and is computed as $S_l(l_1, l_2) = 4l_1l_2/(l_1 + l_2)^2$. $S_a(\triangle abc, \triangle def)$ is the angle similarity between two triangles and is computed by

$$\begin{aligned} S_a(\triangle abc, \triangle def) = 1 - (|\angle bac - \angle edf| + \\ |\angle abc - \angle def| + |\angle acb - \angle dfe|)/360 \end{aligned}$$

We map A_i^1, A_i^2 and A_i^3 to all of their possible candidates, and compute the maximum of relational similarity between A_i and B_{i_k}, which is treated as structural similarity $\tilde{S}(A_i, B_{i_k})$ between A_i and B_{i_k}.

$$\tilde{S}(A_i, B_{i_k}) = \max_{\mathcal{P}} S(A_i, B_{i_k}, \mathcal{P})$$

4. When all structural similarities between regions and their corresponding candidates have been computed, new matching similarity $S(A_i, B_{i_k})$ for A_i and its corresponding candidate B_{i_k} is calculated by normalizing $\tilde{S}(A_i, B_{i_k})$ by $\sum_j \tilde{S}(A_i, B_{i_j})$.

5. All the processes and computations above are also carried out for every region B_j in the other image.

6. We intergrate the matching similarities computed for two images by

$$\hat{S}(A_i, B_j) = (S(A_i, B_j) + S(B_j, A_i))/2 = \hat{S}(B_j, A_i)$$

and reselect the correspondence for each region based on the intergrated matching similarities.

7. Steps 3—6 are iterated until the results of matching do not change.

The matching results of our method for the two images in Fig. 1 are shown in Fig. 2. Although it worked fairly well in this case, due to the robustness of our ternary relations compared with binary relations such as region adjacency, there are still some mistakes. In the following stage we consider how to detect and correct these mistakes using the global consistency of the whole obtained correspondences.

The validating and correcting stage

Based on the assumption that the scene does not change while the camera has moved, if the correspondences are correct, then the locational shifts between the corresponding regions in two images can be regarded as the movements of the objects in the image plane due to the camera movement. It means that these locational shifts should follow the optical flow generated by the camera movement. The locational shifts of the center of gravity of regions obtained from the region correspondences for images in Fig. 1 are shown in Fig. 3. From the figure we can observe that the locational shifts of the correct correspondences are rather regular in the optical flow while those of the incorrect correspondences violate the whole regularity apparently.

Under the hypothesis that most of the obtained correspondences are correct, we can estimate the parameters of the camera movement using all the obtained correspondences. Then we check each correspondence with the estimated parameters of the camera movement and judge a correspondence to be a mistake if its locational shift in two images is far from the one predicted from the estimated parameters.

Let (x_i, y_i) and (x'_i, y'_i) be the center of gravity of a region in one image and the corresponding region in the other image, respectively, and let

$$R = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix}, \text{ and } T = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

be the rotation matrix and the translating vector of movement of the scene determined by the camera movement, then we can get the epipolar condition as follows:

$$\begin{pmatrix} x' & y' & 1 \end{pmatrix} T_\times R \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0$$

where $T_\times R$ is defined as

$$T_\times R = \begin{pmatrix} 0 & -Z & X \\ Z & 0 & -Y \\ -X & Y & 0 \end{pmatrix} R$$

Given a set $\{(x_i, y_i) \to (x'_i, y'_i)\}$, which represents the locational shifts of centers of gravity of the corresponding regions, we can estimate the parameters of

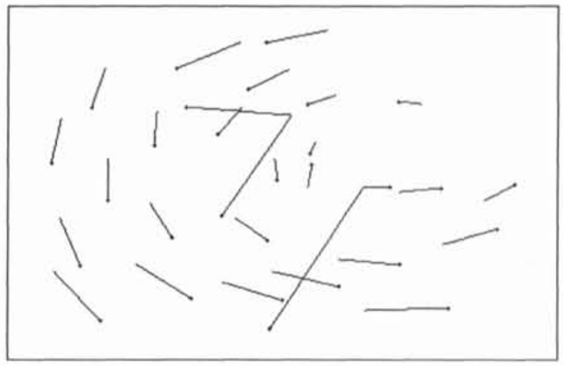

Fig. 3: Locational shifts of regions

the camera movement R and T up to a scalar using the linear method described in [4].

We propose here two ways to check out the outliers of the correspondences which are not coherent with the whole consistence of optical flow determined by the camera motion parameter estimated from all the correspondences.

One way is based on an instinctive fact that if the outliers are few, then parameters estimated by discarding an outlier from the correspondences will perturb much more from parameters estimated with all the correspondences than that estimated by discarding a correct correspondence. Let P_i be the value of the perturbation of the parameters estimated by discarding of the ith correspondence, then $\{P_i\}$ will be bimodal, one consists of small values related to discarding a valid correspondence and the other consists of large values related to discarding an outlier.

Another way is to compute the error in the image plane for every correspondence under the estimated parameters of camera motion. If a point (x, y) in one image has a correspondence (x', y') in the other image, then (x', y') should be on the line segment from

$$\begin{pmatrix} x'_0 \\ y'_0 \end{pmatrix} = \begin{pmatrix} (r_1x + r_2y + r_3)/(r_7x + r_8y + r_9) \\ (r_4x + r_5y + r_6)/(r_7x + r_8y + r_9) \end{pmatrix}$$

to

$$\begin{pmatrix} x'_1 \\ y'_1 \end{pmatrix} = \begin{pmatrix} (x'_0 + X)/(1 + Z) \\ (y'_1 + Y)/(1 + Z) \end{pmatrix}$$

if $Z \geq 0$, or to

$$\begin{pmatrix} x'_1 \\ y'_1 \end{pmatrix} = \begin{pmatrix} x'_0 + X \\ y'_1 + Y \end{pmatrix}$$

otherwise.

Let (x_i, y_i) and (x'_i, y'_i) are the centers of gravity of two regions of a correspondence, (x'_i, y'_i) should lie near the line segment determined by (x_i, y_i), R and T as above. So the error of the correspondence can be computed as the distance from (x'_i, y'_i) to that line segment.

Similar to the distribution of the perturbations of estimated parameters by discarding single correspondences, the error distribution of the correspondences in the image plane is also bimodal. Several statistical methods can be used to find a threshold to separate these two modes. We separate the outliers from the valid correspondences using the thresholding method proposed by Kittler and Illingworth[5].

For each outlier of the correspondences detected above, correction is carried out as follows.

First the estimated optical flow of each correspondence is computed using the estimated parameters of camera movement and the estimated relative depths. For the outliers detected, the depths are substituted by the average depth. The estimated optical flow is shown in Fig. 4.

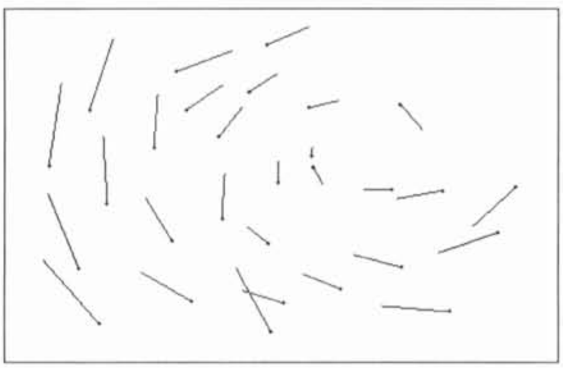

Fig. 4: Estimated optical flow

For a region A_i whose correspondence is judged as an outlier, if there is a corresponding candidate B_{i_k} which has the highest matching similarity with A_i among the ones which are consistent with the estimated optical flow of A_i. We say that B_{i_k} is consistent with the estimated optical flow of A_i if the vector from the center of gravity of A_i to that of B_{i_k} lies in a area from $-10°$ to $10°$ relative to the direction of the estimated optical flow, and the length of the vector is from $\frac{2}{3}$ to $\frac{3}{2}$ of that of the estimated optical flow.

If an outlier of correspondences could not be corrected with the step above, we try to use the ternary relation of regions to correct it. For a region A_i whose correspondence is an outlier, select two nearest regions A_i^1, A_i^2 whose correspondences are valid. By mapping A_i^1 and A_i^2 to their correspondences and the triangle, which consists of the centers of gravity of A_i, A_i^1 and A_i^2, simultaneously onto the other image, we can estimate the approximate location $(\hat{x}_i, \hat{y}_i)$ where the center of gravity of the corresponding region of A_i should be. We find regions which are similar to A_i in color, within a disk with the center locating at $(\hat{x}_i, \hat{y}_i)$ and the radius equal to half the length of the longer of the two edges from $(\hat{x}_i, \hat{y}_i)$ of the mapped triangle. Among these regions, the one have the biggest area in the disk is assigned as the correspondence of A_i.

A region A_i whose correspondence is an outlier which could not be corrected by the above steps is judged to have no correspondence.

The corrected correspondences of Fig. 2 are shown in Fig. 5, where the outliers are checked out with the first method. A '0' means that no corresponding region exists in the other image.

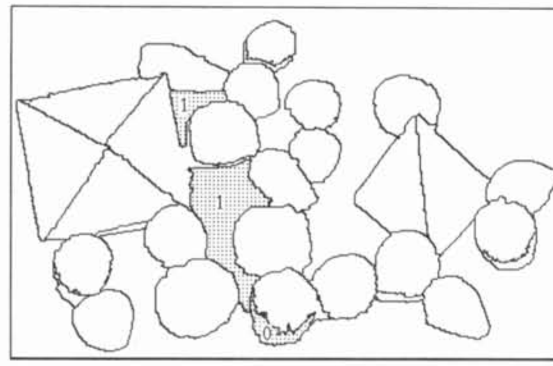

Fig. 5: Corrected correspondences(for left image)

Conclusion

We proposed a region-based approach for image correspondence. Our approach adopted a two-stage matching strategy, that is, a matching stage followed by a validating and correcting stage. Region correspondences are first established using local ternary relations among regions together with region features. The majority of the obtained correspondences are guaranteed to be correct by using the local ternary relations among regions in matching. Then optical flow generated by the camera movement is estimated from all the obtained correspondences and is used to detect and correct some apparent mistakes in the matching results. Experimental results show the effectiveness of our method.

Acknowledgement

This study is partly supported by Real World Computing Partnership and Tokai Foundation for Technology.

References

[1] S. B. Marapane and M. M. Trivedi: Region-Based Stereo Analysis for Robotic Applications, IEEE Trans. on SMC, Vol. 19, No. 6, pp. 1447-1464, 1989

[2] Y. L. Chen: Image Region Correspondence Based on Structural Similarity and Color Information, Doctor Thesis, Sizuoka University, 1993

[3] D. Yagi, K. Abe and H. Nakatani: Segmentation of Color Aerial Photographs Using HSV Color Models, Proc. IAPR Workshop on Machine Vision Applications, pp. 367-370, 1992

[4] J. Weng and T. S. Huang: Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation, IEEE Trans. on PAMI, Vol. 11, No. 5, pp. 451-476, 1989

[5] J. Kittler and J. Illingworth: Minimum Error Thresholding, Pattern Recognition, Vol. 19, pp. 41-47, 1986