Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Combining Color Histogram and Gradient Orientation Histogram for Vision Based Global Localization

Hong Liu
Key Laboratory of Machine Perception and Intelligence,
Key Laboratory of Integrated Microsystem,
Shenzhen Graduate School, Peking University, CHINA
E-mail:hongliu@pku.edu.cn

Xiaojia Yu Haitao Yu
Key Laboratory of Integrated Microsystem,
Key Laboratory of Machine Perception and Intelligence,
Shenzhen Graduate School, Peking University, CHINA
E-mail: yuxj06328@szcie.pku.edu.cn
yuht06130@szcie.pku.edu.cn

***Abstract*— Global image features and local image features are comprehensively used in mobile robot's localization. In this paper, we proposed a geometric approach based on the combination of global image features. Considering the deficiency of the Weighted Gradient Orientation Histograms (WGOHs) for similar structure environments, color histograms are integrated as one vector for the localization. Besides the improving of weight and division for WGOH and color histogram, another weight for different emphasis on color and gradient orientation is carried out. A normalizing process is performed to better integrate the two global features. This combining approach is tested by means of locations recognition. Experimental results show that the proposed combining approach is efficient for indoor environments.**

***Keywords*—Color histogram, WGOH, Global localization**

.

I. INTRODUCTION

For a mobile robot, localization and navigation problems are essential to serve in practical environments. The former is navigation problem, and the latter is localization problem. When a mobile robot moves from a start position to a goal destination, it has to know where it is. Localization is to locate mobile robots' current position at any time.

At present, localization problem usually can be classified as global localization and local tracking [1] [2]. Global localization is to realize robot self-localization with no prior information except that the robot is in the environment. However, local tracking is to locate the robot's position with not only the current information but also the prior information where the robot's initial position and previous position are. Most local tracking methods are realized by probabilistic approaches, such as Markov [3], particle filter [4] and so on. In comparison, global localization is relatively simple.

Environment representation is how the robot represents the practical environments. It is a basic problem for both global localization and local tracking. Globally, it can be categorized into geometric approaches and topological approaches [5]. Geometric approaches typically use a two-dimensional grid as a map representation to keep track of the robot's exact position. While, topological approaches use an adjacency graph to represent environments. Though locations can be located exactly with the geometric approach, the cost will be very expensive. In contrast, topological approach can be implemented simply, rapidly and stably.

In vision-based topological localization, each location is represented by images. Many image features have been successfully used to recognize locations in mobile robots self-localization. Usually, these image features can be categorized into local image features and global image features. A few interesting points, which belong to a special part of the image, are used for local features-based localization. These features generally are invariant to scale and rotation, therefore, local feature based localization approach can be robust to occlusion. However, the distinctive features are very difficult to extract. Global feature-based localization is to use features of the whole image. The computation of global features is rapid, and matching of those features is also convenient. SIFT (Scale Invariant Feature Transform) [6] [7] and Harris [8] are the most famous local features.

Global image features are successfully used for topological global localization in many localization systems. Urich presented color histograms for localization [5]. The six components of R, G, B, H, S, and V were used to form six histograms, and a voting scheme was introduced to realize localization. Multi-dimensional histogram was used for mobile robot localization by T. Tan [9]. J. Kosecka et al. [10] proposed an approach of inferring a topological environment model. Learning Vector Quantization (LVQ) technique was adopted to obtain sparser representation for each location. Then, image gradient orientation was considered as the distinguishing characteristic of each individual location.

Some approaches, which combine global image features and local image features, are applied to mobile robots. Weiss et al. presented a hybrid approach [11]. For most of the images, global image features were used considering the localization speed. While others used local image features, SIFT, for the localization accuracy. Particle filter was used to decide which feature would be used. Wang et al. brought forward a coarse-to-fine vision-based localization approach [12]. LVSM (location vector space model) was built for localization. Harris detector and SIFT descriptor were used for the whole process. Weighted Gradient Orientation Histogram (WGOH) and Weighted Grid Integral Invariants (WGH) were combined for global localization [13]. However, as long as local interest

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

features [3] [11] [12] were extracted, cost was high and localization process was time consuming.

TABLE I. THE COMPARATIVE RESULTS OF LOCALIZATION APPROACHES

Approaches	*Merits*	*Disadvantage*	*Works*	*Methods*
Global image features	Compute rapidly and conveniently	The localization accuracy is not high.	I.Urich et al.	Color, texture, gradient
Local image features	Invariant to scale and rotation, robust to occlusion, high localization accuracy	Compute expensively, compare inconvenient.	Lowe et al.	SIFT, Harris feature
Hybrid approach	High localization accuracy, robust to occlusion.	The computing cost is still huge.	Weiss et al.	Combining the WGOH and SIFT

In some environments, it is difficult to realize localization by using the color property, especially for changing illumination condition. Gradient orientation is a feasible approach to distinguish different images. However, it is difficult to distinguish locations when scenes are very similar and gradient orientations are alike. Color is an effective complementary feature to gradient orientation.

In this paper, weighted color histograms and weighted gradient orientation histograms are combined for similar structure environments. Division and weight are improved on color histograms and gradient orientation histograms. An approach similar to that described in this paper was used for multimedia retrieval by Michael Ortega [15]. But, the task of place recognition is more complex than that in image retrieval for similarity among locations.

The rest of this paper is organized as follows. Section Ⅱ describes the image feature extraction. The weighted color histogram and WGOH are presented, respectively. In section Ⅲ, how two histograms are combined is introduced. Section Ⅳ lays out our experiments' results, and in section Ⅴ, conclusions of this paper are made.

II. IMAGE FEATURE EXTRACTION

The environments are represented into topological form in current work. The camera of the robot captures images of the environment from each location. Then locations are related with images. Localization task becomes to find the most similar images among all location's images. The localization process can be divided into two stages, processing stage and localization stage. In processing stage, standard images are related with locations during the robot's moving. And, images are processed to form a database with an approach, which combines color histogram and WGOH. In localization stage, once the robot gets a location, localization results are obtained through comparing with images in database.

Color histogram is processed with weighting and division for better results. Also, color histogram and WGOH are combined as the image feature through an efficient normalization process.

A. Weighted Color Histogram

Color is a very common but important feature in many fields. Also, color histogram is used by a lot of image matching systems. The details of color histogram were described by M. J. Swain [14]. The most distinctive characteristic of color histogram is that it can both efficiently express the image features and be computed quickly. It only needs very little memory to process the histograms. To improve the performance of classical color histograms, the division is carried out, which is shown in Fig 1.

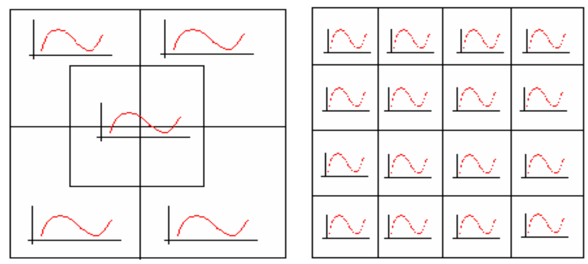

Figure 1. The divided color histogram (left) and WGOH (right)

HSV color space is used for localization. Compared to RGB space, the three components of HSV space are more stable. And, our experiments have proved its efficiency. After RGB space transformed into HSV space, quantization is carried on to each component. As H component plays a more important role, it is quantified into 16 bins. S component is quantified into 4 bins. Since V component is easily affected by illumination conditions, only HS components are considered. Then, HS components make up a 20 bins' vector.

The disadvantage of color histograms is that the relationship between pixels is not considered. Two images may share similar histograms; however, contents are quite different. The images are divided into 5 sub-images to improve performances. Firstly, the image is divided into 2×2 sub-images to form 4 sub-images. And another sub-image is got through the center of the image, which is the same size as other sub-images. Then, histogram of each sub-image is built, and all sub-images' histograms are stacked to form one vector. After the process of each image, a 20×5 vector is constructed. All reference images are processed, and corresponding vectors make up the reference database.

Besides, color histogram is weighted. Classical histograms don't consider space position of pixels. Therefore, position relationship between pixels is used to weight histogram. The pixel is weighted by the distance to the center of the sub-image. Also, this weighting approach can deal with small translations of positions well. Color histograms of two locations have been shown in Fig. 2.

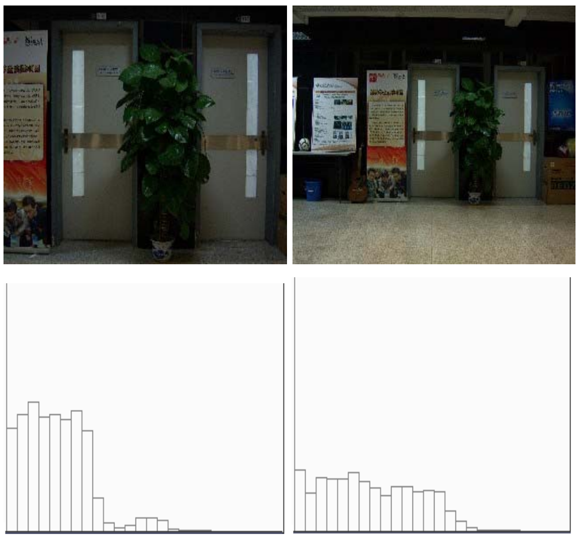

Figure 2. Scenes and their associated color histograms

B. Weighted Gradient Orientation Histograms

Weighted gradient orientation histogram (WGOH) is very popular in both outdoor localization and indoor localization. This method was presented by Bradley et al. [16]. Several improving and simplifies has carried out on it. Gradient orientation can well deal with illumination changes. Several systems have proved that this method is an efficient approach and can make good recognition results.

Unlike color histogram, WGOH first converts an image into gray image. All the processes are based on the gray-image. Then gradient magnitude $m(x, y)$ and gradient orientation $\theta(x, y)$ of each pixel are computed by the formulas (1) and (2), respectively,

$$m(x,y)=\sqrt{\left(G(x+1,y)-G(x-1,y)\right)^2+\left(G(x,y+1)-G(x,y-1)\right)^2} \quad (1)$$

$$\theta(x,y)=\begin{cases} \pi/2 & G(x+1,y)=G(x-1,y) \\ 0 & G(x,y+1)=G(x,,y-1) \\ \arctan\left(\dfrac{G(x,y+1)-G(x,y-1)}{G(x+1,y)-G(x-1,y)}\right) & others \end{cases} \quad (2)$$

here, $G(x, y)$ is the gray value of each pixel. The computation of contour pixels is through appointing outside pixels' values same to contour pixels'.

Gradient magnitude is normalized to better weight orientation. Firstly, the maximum of the whole image's gradient magnitude is computed as *Max*. Then, normalizing is realized by the formula (3).

$$\overline{m(x,y)}=\frac{m(x,y)}{Max} \quad (3)$$

where, $\overline{m(x, y)}$ is the normalized gradient magnitude of pixel.

On one hand, gradient orientation is weighted by the normalized gradient magnitude. Gradient magnitude is selected to weight orientation, because it increases relativity of pixels. On the other hand, images are divided into 4×4 sub-images. And, orientation is also weighted by the distance to the center of the sub-image. Besides, all sub-image's orientation histograms are stacked together to form one vector. Since orientation is quantified into 8 bins, a 16×8 final orientation vector is formed.

III. COMBINING WCH AND WGOH

A. Feature Normalization

Unlike Weiss et al. [13], which combined WGOH and WGH with only simple multiplication of two features, a normalizing process of features is used to combine the color histograms and WGOH features.

Algorithm 1: Combining Color Histogram and WGOH

Preprocessing phase

Required: each image feature: H_r, r = 0,…,M-1, and weighted coefficients: ω_1, ω_2,

1: weight H_r

$$\begin{cases} H_{ri}=\omega_1\cdot H_{ri} & (i=0,\ldots,L-1) \\ H_{ri}=\omega_2\cdot H_{ri} & (i=L,\ldots,N) \end{cases}$$

2: compute the similarity between each features $D(H_i, H_j)$:
$D(H_i, H_j)$ = dist(H_i, H_j)
i, j = 0…M-1, $i \neq j$

3: set D_i={each distances of database } i = 0,…,(M-1)*M/2

4: **for** each $d \in D$
Compute the mean: $m = \sum d_i/((M-1)*M/2)$

5: **for** each $d \in D$
Compute the deviation: $\sigma = \sum (d_i-s)^2/((M-1)*M/2)$

Localization phase

Required: input feature: H

1: weight H

$$\begin{cases} H_i=\omega_1\cdot H_i & (i=0,\ldots,L-1) \\ H_i=\omega_2\cdot H_i & (i=L,\ldots,N) \end{cases}$$

2: compute the similarity between input feature and reference features $s_i(H, H_i)$, i = 0,…,M-1

3: set $S = \{s_0, s_1, \ldots, s_{M-1}\}$

4: **for** each $s_i \in S$
Normalize $s_i` = ((s_i-m)/(3*\sigma)+1)/2$

5: compute the minimum of $s_i`$

Figure 3. Combining Color Histogram and WGOH for Localization

Usually, each element of a feature makes equal contribution to the whole vector. For example, a vector of color histogram makes equal emphasis of each element. All the elements represent the numbers of the color components. However, if the elements represent different signification of the vector, it needs to be processed to make no differences. Considering the sequences {0.1, 0.5, 0.7, 100, 200}, it is obvious that the elements make different senses. Also, it needs to fairly treat the comparing distances when matching images is carried out. The normalization procedure can realize the above purposes well.

Formula (4) is used to realize the normalization process for its advantages mentioned in [15]. Through this process, about *99%* feature elements are constrained into [-1, 1]. And convert it into [0, 1] through formula (5).

$$s = \frac{s_i - m}{3\sigma} \tag{4}$$

$$s' = \frac{s+1}{2} \quad . \tag{5}$$

where, s_i is responding element of all features, m and σ are mean and deviation of elements, respectively.

In preprocessing phase, features of the whole image database are obtained through color histogram and WGOH processing. Firstly, weighting is carried out to make color and gradient orientation have different proportion. Secondly, distances are computed to all the other database features. At last, mean and deviation of all the distances is calculated and saved. The number of distances is *(M-1)*M/2*, where *M* is the feature numbers.

In localization phase, the feature of the new image is also processed by color histogram and WGOH. At first, the feature is weighted for color components and gradient orientation components respectively. Then, similarities with all the database features are computed. The normalization process is put up for each similarity. At last, the minimum similarity is chosen as the robot's current location. The details of the process are shown in Fig 3. *M* is the number of database images. *L* and *N* is the number of color elements and gradient orientation elements, respectively.

B. *Matching approach*

Many matching approaches for distances computing have been tried in the experiments, including histogram intersection, χ^2distribution, Jeffrey divergence and so on [17]. The histogram intersection is used for its simplicity and efficiency. The definition of histogram intersection is shown as:

$$D(h_i, h_r) = \frac{\sum \min(h_i, h_r)}{\sum h_i} . \tag{6}$$

where, h_i is the histogram of input image, and h_r is the histogram of reference image.

C. *Weighting color and gradient orientation*

To deal with the relationship between two features better, weighting color and gradient orientation is carried out. The above-metioned weighting is to increase the relativity of pixels, which reflect different contributions to the feature.

Another different weighting is carried out, which is shown as formula (7). Color weight and gradient orientation weight are put into different value for different scene. This reflects different emphasis on the color and gradient orientation to the scenes. The different weighting expression can give more freedom to the user. However, an auto-adaptive weighting approach is expected.

$$\begin{cases} \omega_1 + \omega_2 = 1 \\ f_i = \omega_1 \cdot f_i & (i = 0, \ldots, N-1) \\ f_i = \omega_2 \cdot f_i & (i = N, \ldots, M) \end{cases} \tag{7}$$

here, ω_1 and ω_2 are the weighting coefficient. f_i is the element *i*of the feature. *N* is the number of the color elements, and *M* is the number of the element of the feature.

IV. EXPERIMENTS AND ANALYSES

Experiments were carried out on our lab's robot in indoor environments. In order to control locations better, the robot was pulled manually. Images were captured by still camera on the robot. The size of each image is 320×256, as this is enough for our localization. Unlike other's robot localization systems, on one hand, the localization precision is more accurate; on the other hand, the speed of localization is more rapid to meet the requirements of real-time. The precision of some systems usually are 2 meters or longer than 2 meters between two locations [9]. However, the interval between two neighboring locations of our experiments is only 60 centimeters, as this can make the robot locate itself more accurate.

If the localization results meet the following conditions, the localization result is considered to be successful. Firstly, the location is really the nearest to the located position among all locations, which have been saved in the database. Secondly, if several standard locations are the same far away the location and one of them is located, the result is still right.

Two scenes are tested during different time in our experiments. One hundred standard images were captured as reference images. Other different position's images to be located are radomly selected, but not consistent with standard locations. The structure of this scene is very similar, like many grid-like windows. GPS wasn't used in our experiments, and all the positions were labeled manually. Methods of machine learning are planned to carried out in future work. Each of the position is marked and ensured for several times. The camera's orientation of each position is coherent, but a certain degree errors are inevitable.

TABLE II. COMPARATIVE RESULT S OF APPROACHES IN OUR ENVIRONMENTS

Approaches	**The Successful Localization Images**			
	Successful Rate	***Type 1***	***Type 2***	***Type 3***
Color Histograms	85.45%	√	×	×
WGOH	67.27%	×	√	×
Combining Approach	90.91%	3.64%	5.45%	3.64%
CH • WGOH	81.82%			

The approach of simple multiply of two vectors like Weiss's was tested and compared with our method. Color histogram localization, WGOH localization and the combining of two approaches has been tested respectively.

In TAB Ⅱ, CH • WGOH represents the approach, which was used by Weiss [13]. Type 1 represents correct results of using combining approach, while the color histogram approach is right and WGOH approach is wrong. Type 2 represents the correct results using our combing approach, while the color histogram is wrong and WGOH is right. Type 3 is the right result of using our combining approach; however, the result of using the other two approaches is both wrong. Our combining approach was improved by weighting WGOH and color with the rate 0.25. In the wrong results, 5.45% of them are the same as WGOH, and the other 3.64% results are the same as color. But there are another 3.64% results successfully located, which were located wrongly both by the color and WGOH. An example is showed in Fig 4.

Figure 4. Example of an image from a location (left) and located image (right). The above is the correct result using the combining approach, and the below is wrong using other three approaches respectively.

For the similarity of the environments, WGOH isn't the best matching approach for localization. However, the color is more stable. Several wrong images, which were located by the WGOH, were successfully located by the color feature. By the weighting of color and WGOH, the best result can be chose. These two approaches are complementary, and combining them together can make localization more accurate. Therefore, our combining approach is an efficient approach for similar environments.

V. CONCLUSIONS

Considering the accuracy of localization, most methods use local image features, like SIFT. However, the speed of localization usually can't be satisfactory for the expensive computation of local image features. Considering the advantage and disadvantage of color and gradient orientation, classical color histogram and the popular WGOH are combined for similar structure environments localization. A normalization process is successfully applied for the combination.

ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China (NSFC, No.60675025 and No.60975050) and National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247), Shenzhen Bureau of Science Technology and Information.

REFERENCES

[1] G. N. DeSouza, and A. C. Kak, "Vision for mobile robot navigation: a survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 2, Feb. 2002, pp. 237-267.

[2] B. L. Boada, C. Castejon, and L. Moreno, "Topo-geometric modeling and localization in indoor environments," In the 28th Annual Conference of the IECON, Vol. 4, Nov. 2002, pp. 2587-2592.

[3] J. Kosecka and F. Li, "Vision based topological markov localization," In: Proc.ICRA'04, Vol. 2, April 2004, pp. 1481-1486.

[4] C. Weiss, A. Masselli and A. Zell, "Fast vision-based localization for outdoor robots using a combination of global image features," in Proceedings of the 6th Symposium on Intelligent Autonomous Vehicles (LAV 2007), Toulouse, France, September 3-5, 2007.

[5] I. Ulrich and I. Nourbakhsh, Appearance-based place recognition for topological localization, In Proceedings of the IEEE International Conference on Robotics & Automation (ICRA), San Francisco, April 2000, pp 1023-1029.

[6] D. Lowe, "Distinctive image features from scale-invariant keypoints," Int. Journal of Computer Vision, Vol. 60, No. 2, 2004, pp. 91-110.

[7] S. Se, D. Lowe, and J. Little, "Local and global localization for mobile robots using visual landmarks," In Proc. IROS'01, Maui, Hawaii, Oct. 2001, pp. 414-420.

[8] C. Shmid and R. Mohr, "Local grayvalue invariant for image retrieval," IEEE Trans. Pattern Anal. Mach. Intell., Vol. 19, No. 5, May 1997, pp. 530-534.

[9] C. Zhou, Y. Wei, and T. Tan, "Mobile robot self-localization based on global visual appearance features," In Proc. ICRA'03, Vol. 1, Sept. 2003, pp. 1271-1276.

[10] J. Kosecka, Zhou Liang, P. Barber and Z. Duric, "Qualitative image based localization in indoors environments," In: Proc. CVPR'03, Vol. 2, June 2003, pp. 3- 8.

[11] C. Weiss, H. Tamimi, A. Masselli and A. Zell, "A hybrid approach for vision-based outdoor robot localization using global and local features," In Proc. IROS'07, San Diego, CA, USA, pp. 1.47-1052.

[12] J. Wang, H. Zha, and R. Cipolla, "Coarse-to-fine vision-based localization by indexing scale-invariant features," In IEEE Transactions on Systems, Man, and Cybernetics, Vol. 36, No. 2, April 2006, pp. 413-422.

[13] C. Weiss, A. Masselli, and A. Zell, "Fast vision-based localization for outdoor robots using a combination of global image features," In Proc. 6th Symposium on Intelligent Autonomous Vehicles (IAV 2007), Toulouse, France, September 3-5, 2007.

[14] M. J. Swain and D. H. Ballard, "Color indexing," International Journal of Computer Vision, Vol. 7, No. 1, 1991, pp. 11-33.

[15] M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, and T. S. Huang, "Supporting Similarity Queries in MARS", in Proc. of ACM Conf. on Multimedia, 1997, pp. 403-413.

[16] D. M. Bradley, R. Patel, N. Vandapel and S. M. Thayer, "Real-time image-based topological localization in large outdoor environments," In: Proc.IROS'05, Edmonton, Canada, pp.3062-3069.

[17] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," International Journal of Computer Vision, Vol. 40, No. 2, 2000, pp. 99-121.