Abstract — In Mobile Robotics, the Initial Global Localization Problem is currently a major challenge. The visual approach to solve this problem is very sensitive to the number and quality of information presented in captured images. Thus, this paper evaluates the influence of image pre-processing techniques on the proposed appearance-based characterization model. In this paper we evaluated 4 image processing filters: gaussian, weinerner, bilateral and kuwahara, aiming to get a good cost-benefit index between outliers filtering and reduction in the number of features used to characterize the places. This model used SURF-128 features and is based on probability mass functions (PMF). The experiments were executed in an outdoor environment to evaluate the characterization power of the model and the effectiveness of the 4 filters applied. After pre-filter the images with Wienner filter it was achieved 77% in AUC parameter from the global ROC curve and reduction of 13% in the number of originally extracted features from the places.

Keywords — Initial Global Localization, Probabilistic Model, Image Pre-filtering, Localization and Mapping.

1 INTRODUCTION

Mobile Robotics has many problems that address the issue “where is the robot right now?”. These problems are defined and compared in Cupec et al. (2015) and their generalization is called the Initial Global Localization Problem, that can be defined as the determination of the robot location in an environment assuming no prior information.

The visual solution for that problem has two main approaches: appearance-based and feature-based (Cupec et al., 2015). Feature-based approach uses geometric features, that can be computed using metric information such as 3D reconstructed points, lines and other geometric elements.

On the other hand, in appearance-based approach, each location in the robot’s operating environment is defined by descriptors extracted from it. Then, the localization is performed by matching the current image’s descriptors and the descriptors assigned to each place in the environment.

In this work, we present a modification in the appearance-based place recognition model, firstly presented in de Oliveira et al. (2011), and discuss the effect of pre-filter the images. The new model can achieve better localization performance compared to its former version, and the pre-filtering can reduce the number of features extracted from the places. Thus, we discuss the relationship between classification power and computational efficiency.

1.1 Related work

Considering the first methods of visual features extraction based on interesting points, Lowe and Little (2001) and Andreasson and Duckett (2004), used Scale-invariant Feature Transform (SIFT) (Lowe, 1999) to environment feature extraction. These features were used as landmarks in the localization algorithm of a robot. This method incorporates robustness to the localization process by using features invariant to rotation, translation and illumination. As a drawback, such developed application was tested mainly in indoor environments and not outdoors, where more significant changes in the environment usually occur.

Herbert Bay and colleagues proposed in Bay et al. (2006) and Bay et al. (2008), the Speeded Up Robust Features (SURF) method, that outperformed the cost-benefit index of the SIFT. SURF presents excellent results in relation to accuracy and a feasible computational cost, that allows applications in online autonomous systems.

Currently, considering the appearance-based approach for localization, one of the most significant contributions is the algorithm Fast Appearance-Based Mapping (FAB-MAP) (Cummins and Newman, 2008) and its further improvement FAB-MAP 2.0 (Cummins and Newman, 2010). Such work is a new formulation of Simultaneous Localization and Mapping (SLAM), and it was very successful because of the tests performed over large distances.

An important contribution of the paper Cummins and Newman (2010) was the execution of two tests in an outdoor environment: one of them was a 7 km route and the other a 1000 km route. Images had been captured on these routes and the generated database was used to evaluate the model.

Despite both tests had been successful, the algorithm had to be trained offline. The training step used 1921 images that produced 2.5 millions of SURF descriptors with 128 elements in the neighborhood that occupied 3.2 GB on memory.

It is important to notice that FAB-MAP can
use primer information to estimate the current location and the images used for testing were captured during only six days. This characterize low chances of having a significant variation in the environment during the tests, which may benefit the FAB-MAB performance.

Also, Mitsuhashi and colleagues, in Mitsuhashi and Kuroda (2011), developed a system for mobile robots localization, using appearance-based place recognition with SURF descriptors. All tests were made in outdoor environments and the average position error obtained was less than 2% with no usage of GPS. However, the obtained results were achieved by sensors fusion and by the utilization of FAB-MAP algorithm (Cummins and Newman, 2008).

2 LOCALIZATION MODEL

The model proposed in this paper used a probabilistic approach to solve the initial global localization problem.

Each place in the environment is represented by a probability mass function (PMF) of SURF-128 features, estimated by a histogram. The features are extracted from \(n\) complete samples\(^1\) of the place.

This model differs from bag-of-words mainly because there is no need for choosing a common vocabulary to represent the classes, places in this case. Each place \(l_i\) is represented by the occurrence of its different features \(F_i\), extracted from the images used to represent it. Then, each place has a local vocabulary \(F_i\) and a local originally extracted vocabulary \(FO_i\). Thus, the whole environment has a set of local vocabularies \(F\) and a set of local originally extracted vocabularies \(FO\).

Another difference from bag-of-words is that it is not computed the occurrence of features in each image to calculate the features’ histogram. Occurrences are computed along \(n\) complete samples to do estimate the probability mass function of features from a place, \(pmf(F_i)\).

In de Oliveira et al. (2011), it was extracted SURF-128 features around a place using two images with approximately 180 degrees field of view, performing a 360 degrees place sample. However, it is important to point that it is not necessary to use images with an 180 field of view. It is possible to use images with a smaller field of view, nevertheless it will be necessary to take more than two images to compose a complete sample.

\(^1\)A set of images that represents entirely a place. In an outdoor environment a complete sample was considered as a combination of images that represents 360 degrees around a place.

\(^2\)This original vocabulary uses each extracted feature as different from the others without evaluation.

2.1 Initial Place Characterization

A traditional Naive Bayes classification (Lewis, 1998), uses the maximum a posteriori probability in (1), for place \(l_i\) and a non-localized feature set \(S\), called here a place sample.

\[
p(l_i|S) = \frac{p(S|l_i)}{p(S)}. \tag{1}
\]

Considering the global localization problem, \(p(l_i)\) has equal probability along the places,

\[
p(l_i|S) = \frac{p(S|l_i)}{p(S)}. \tag{2}
\]

However, the proposed model do not limit features to a previously defined global vocabulary \(F\). It was made this choice because outdoor places are dynamic environments and then their visual characteristics varies depending on light conditions, time of the year and human activities.

So, each place \(l_i\) has its own local features \(F_i\). Then, the bayesian approach is reduced to the maximum likelihood (ML) approach. Thus, (2) is simplified by

\[
p(l_i|S) = p(S|l_i), \tag{3}
\]

implying that features space is locally defined, so each place has its own feature vocabulary \(F_i\). The definition of the vocabulary is made automatically by defining a set of images to characterize the place.

To calculate \(p(S|l_i)\) is necessary to define the probability mass function \(pmf()\) of features \(F_i\) in place \(l_i\). Each place is represented by a PMF of its local features extracted from its characterization image data set.

The ML approach in equation 3 can be justified by computational efficiency. If it was initially defined a global vocabulary of features \(F\), for every new place added in the robot’s operating environment, it would be necessary to redefine the global vocabulary, and consequently to perform matching or clustering algorithms along the features of the entire environment.

2.2 Place Recognition

Once computed the PMF estimation of a place, the goal is to calculate the probability that a non-localized complete sample belongs to this known place.

To define the local vocabulary \(F_i\) of a place \(l_i\), we used the original implementation of SURF extraction and matching algorithms (Bay et al., 2008). So, the independence of features in a place is guaranteed by executing the matching algorithm along the features in the local originally extracted vocabulary \(OF_i\), and then to generate \(F_i\).
Thus, the probability that a non-localized complete sample \( S \) belongs to a place \( l_i \) is calculated by the sum of probabilities that every feature \( s_j \) from the sample \( S \), belongs to the place \( l_i \),

\[
p(l_i|S) = p(S|l_i) = \sum_{j=1}^{m} pmf(F_i = s_j), \forall s_j \in S.
\]

(4)

Then, considering a set of characterized places \( L \) in the mapped environment, the localization \( x(S) \), i.e. place recognition, could be performed by

\[
x(S) = \arg\max_i p(l_i|S), \forall l_i \in L.
\]

(5)

Nevertheless, because of the ML approach in (3) it is not consistent to use a traditional Naive Bayes classification. To use an \( \arg\max \) approach it is considered that there is a global equalization in the probabilities calculation, that is not what happens in this case.

So, the proposed localization method compares \( p(l_i|S) \) to a threshold value \( t_i \) and performs a binary localization for place \( l_i \) as follows,

\[
x(S) = \begin{cases} 
-\neg x(l_i) & \text{if } p(l_i|S) < t_i \\
x(l_i) & \text{if } p(l_i|S) \geq t_i.
\end{cases}
\]

(6)

To define those thresholds we used ROC curve analysis because it is usually considered an accurate solution (Sanchez-Gonzalez et al., 2012) for this type of problem. The chosen criterion was the calculation of the best point in the ROC curve that considers equally the true positive rate (TPR) and the false positive rate (FPR).

To create the ROC curves, we calculated the probability that a set of complete samples belongs to each place in the environment. It is important to notice that the set used to define the threshold has to be different from the characterization set.

Then, it is proposed a localization method based on local thresholds \( t_i \) for the set of thresholds in the environment \( T \). The localization algorithm is presented in algorithm 1.

For the localization method represented by algorithm 1, after computing a vector of possible places for localization, it is necessary to choose one of them to determine the location of the sample \( S \). So, function \( \text{bestLocation} \) can be a simple random choice or a customized algorithm.

### 3 HOMEMADE DATA SET

It was made a set of 28 different places\(^3\), characterized by complete samples. Some of them have

\(^3\)Characterization images from 28 places were captured, but the place L18 was not considered in this experiment because it became unreachable due to building constructions around it.

Thus, a binary localization for place \( L \) in the mapped environment, the localization \( x(S) \), could be performed by

\[
x(S) = \arg\max_i p(l_i|S), \forall l_i \in L.
\]

In order to evaluate the system robustness among different conditions, such as changes on illumination and place aspects, it was used a time interval of, at least, one month and different times of the day to capture the images for characterization and test of the localization method, as can be seen in Table 1.

However, to simplify the analysis, in this paper we executed experiments only for the first 10 places from the environment.

Figure 1: The test environment, from which 10 of their 28 places were used to carry out the tests. All places are marked with yellow pins.
Table 1: Time interval between capture of characterization and test images. In the second and third columns, the format is hour:minutes for the beginning and finishing of the images capture process. The fourth column presents the minimum time interval (in months) between images capture for characterization and test.

<table>
<thead>
<tr>
<th>Place</th>
<th>Capture of characterization images</th>
<th>Capture of test images</th>
<th>Minimum interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0</td>
<td>11:00 - 11:05</td>
<td>12:51 - 12:56</td>
<td>5</td>
</tr>
<tr>
<td>L1</td>
<td>17:22 - 17:25</td>
<td>12:06 - 12:09</td>
<td>1</td>
</tr>
<tr>
<td>L3</td>
<td>11:38 - 11:40</td>
<td>16:03 - 16:06</td>
<td>6</td>
</tr>
<tr>
<td>L4</td>
<td>17:09 - 17:12</td>
<td>11:44 - 11:48</td>
<td>2</td>
</tr>
<tr>
<td>L5</td>
<td>16:53 - 16:56</td>
<td>13:54 - 13:59</td>
<td>1</td>
</tr>
<tr>
<td>L6</td>
<td>17:13 - 17:16</td>
<td>14:47 - 14:50</td>
<td>1</td>
</tr>
<tr>
<td>L7</td>
<td>11:52 - 11:57</td>
<td>15:36 - 15:38</td>
<td>6</td>
</tr>
<tr>
<td>L8</td>
<td>11:21 - 11:24</td>
<td>16:52 - 16:59</td>
<td>6</td>
</tr>
<tr>
<td>L9</td>
<td>17:48 - 17:50</td>
<td>14:50 - 14:55</td>
<td>6</td>
</tr>
</tbody>
</table>

The results of the model on this data set will be presented in section 5.

4 IMPROVEMENTS IN THE MODEL: C-IGL

Considering the model presented in section 2, experiments showed a relationship between the quality of localization and the number of features in a place. For most of the places, the greater the number of features in a place, the worse the localization performance on that place.

That result can be explained by the ML approach in (3). This step discards the existence of similar features in different places, that makes the localization largely sensitive to the number of features in a place.

To calculate $P(S|l_i)$ in (4), it is computed the absolute occurrence $fa(l)$ of $F_i$ in $l_i$, and so this value is normalized by the sum of features absolute occurrence $\#(FO_i)$ in the characterization sample set of place $l_i$,

$$p(S|l_i) = \sum_{j=1}^{m} \frac{fa(F_i = s_j)}{\#(FO_i)} \forall s_j \in S \quad (7)$$

Because of that, places with larger number of features will have localization performance degraded by its normalization factor if compared to a place with a small features’ vocabulary.

Analysing the bayesian approach in (1), one can understand the necessity for the division by $p(S)$.

That term is needed to balance the posteriori probability in relation to the occurrence of features in the environment. Different features with the same term $p(S|l_i)$ in (1) can generate different $p(l_i|S)$ depending on its occurrence over an environment defined by $L$.

For example, considering (1), a common set of features $S$ having many occurrences in $L_i$ will generate smaller value of $p(l_i|S)$ than a rare set of feature.

However, the proposed model in (3) do not uses the global approach of a complete bayesian model. Then, to improve the proposed model without calculate $p(S)$, aiming computational performance, an alternative is to define a blind balancing equation.

Instead of calculating the global occurrence of features, it were incorporated to the initial model a term for balancing it in relation to the number of features originally extracted in the place $\#(FO_i)$,

$$p(l_i|S) = \frac{\#(FO_i)}{\sum_{i=1}^{n} \#(FO_i)} p(S|l_i), \forall FO_i \in FO. \quad (8)$$

Simplifying (8) by (7) we have,

$$p(l_i|S) = \frac{\sum_{j=1}^{m} fa(F_i = s_j)}{\sum_{i=1}^{n} \#(FO_i)}, \quad (9)$$
(a) Place L0.

(b) Place L1.

(c) Place L2.

(d) Place L3.

(e) Place L4.

Figure 3: Five image samples for testing the localization model.

the new model called C-IGL.

5 RESULTS

All the experiments performed in this paper used the data set from section 3. Characterization and test samples were built using 4 images to guarantee complete characterization. It was necessary 4 images because the camera’s field of view is smaller than 180 degrees.

The characterization sample set of each place was composed by 10 complete samples, i.e. 10x4 images. The test sample set of each place was composed by 12 complete samples, i.e. 12x4 images. We used window size of 5x5 for all the tested filters.

To test the proposed model it was calculated the probability that a test sample belongs to every place in the environment. This procedure was done using samples from each one of the 10 places. To evaluate the pre-filtering step we used as references the AUC for the original images as much as the number of originally extracted features of them. Tables 2 and 3 present the results for each place (local) and for the environment (global).

The rates Relative Feature Difference (RFD) and Relative AUC Difference (RAD), are calculated by the relative difference, respectively, from the number of originally extracted features and the AUC parameter of the model, without executing pre-filtering step.

One can note, by the negative signals from RFD columns, that all the filters reduced the number of features. Observing the positive signals from RAD columns, one can conclude that the filters Gaussian (G), Wiener (W) and Kuwahara

Table 2: Relative Feature Difference (RFD) x Relative AUC Difference (RAD).

<table>
<thead>
<tr>
<th>Place</th>
<th>Gaussian RFD</th>
<th>Gaussian RAD</th>
<th>Bilateral RFD</th>
<th>Bilateral RAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0</td>
<td>-0.07</td>
<td>-0.16</td>
<td>-0.12</td>
<td>-0.11</td>
</tr>
<tr>
<td>L1</td>
<td>-0.13</td>
<td>0.00</td>
<td>-0.21</td>
<td>0.00</td>
</tr>
<tr>
<td>L2</td>
<td>-0.21</td>
<td>0.00</td>
<td>-0.25</td>
<td>-0.01</td>
</tr>
<tr>
<td>L3</td>
<td>-0.07</td>
<td>-0.02</td>
<td>-0.13</td>
<td>-0.03</td>
</tr>
<tr>
<td>L4</td>
<td>-0.10</td>
<td>0.00</td>
<td>-0.16</td>
<td>-0.03</td>
</tr>
<tr>
<td>L5</td>
<td>-0.09</td>
<td>0.00</td>
<td>-0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>L6</td>
<td>-0.09</td>
<td>-0.23</td>
<td>-0.15</td>
<td>-0.41</td>
</tr>
<tr>
<td>L7</td>
<td>-0.10</td>
<td>-0.28</td>
<td>-0.14</td>
<td>-0.18</td>
</tr>
<tr>
<td>L8</td>
<td>-0.08</td>
<td>0.26</td>
<td>-0.16</td>
<td>-0.06</td>
</tr>
<tr>
<td>L9</td>
<td>-0.11</td>
<td>0.05</td>
<td>-0.17</td>
<td>-0.04</td>
</tr>
<tr>
<td>Global</td>
<td>-0.10</td>
<td>-0.03</td>
<td>-0.16</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

Table 3: Relative Feature Difference (RFD) x Relative AUC Difference (RAD).

<table>
<thead>
<tr>
<th>Place</th>
<th>Wiener RFD</th>
<th>Wiener RAD</th>
<th>Kuwahara RFD</th>
<th>Kuwahara RAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0</td>
<td>-0.11</td>
<td>0.00</td>
<td>-0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>L1</td>
<td>-0.13</td>
<td>0.00</td>
<td>-0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>L2</td>
<td>-0.20</td>
<td>0.00</td>
<td>-0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>L3</td>
<td>-0.10</td>
<td>0.02</td>
<td>-0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>L4</td>
<td>-0.12</td>
<td>-0.16</td>
<td>-0.07</td>
<td>-0.19</td>
</tr>
<tr>
<td>L5</td>
<td>-0.12</td>
<td>0.00</td>
<td>-0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>L6</td>
<td>-0.12</td>
<td>-0.04</td>
<td>-0.07</td>
<td>-0.30</td>
</tr>
<tr>
<td>L7</td>
<td>-0.12</td>
<td>-0.32</td>
<td>-0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>L8</td>
<td>-0.12</td>
<td>0.06</td>
<td>-0.10</td>
<td>0.44</td>
</tr>
<tr>
<td>L9</td>
<td>-0.13</td>
<td>0.00</td>
<td>-0.07</td>
<td>-0.19</td>
</tr>
<tr>
<td>Global</td>
<td>-0.13</td>
<td>-0.03</td>
<td>-0.08</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

ROC curve for all the 10 places. Finally, to compare the model’s performance, we used as a parameter the area under the ROC curves (AUC).

The effectiveness of the improvement from (8) can be confirmed by the reached increment of 11.71% in the global AUC parameter without filter the images, from 68.47% to 80.18%.

To evaluate the pre-filtering step we used as references the AUC for the original images as much as the number of originally extracted features of them. Tables 2 and 3 present the results for each place (local) and for the environment (global).

The rates Relative Feature Difference (RFD) and Relative AUC Difference (RAD), are calculated by the relative difference, respectively, from the number of originally extracted features and the AUC parameter of the model, without executing pre-filtering step.

One can note, by the negative signals from RFD columns, that all the filters reduced the number of features. Observing the positive signals from RAD columns, one can conclude that the filters Gaussian (G), Wiener (W) and Kuwahara...
Table 4: Model’s sensitivity to filters.

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>B</th>
<th>W</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0</td>
<td>2.29</td>
<td>0.92</td>
<td>0.04</td>
<td>-0.62</td>
</tr>
<tr>
<td>L1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>L2</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>L3</td>
<td>0.34</td>
<td>0.21</td>
<td>-0.18</td>
<td>-0.51</td>
</tr>
<tr>
<td>L4</td>
<td>0.01</td>
<td>0.19</td>
<td>1.25</td>
<td>2.78</td>
</tr>
<tr>
<td>L5</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>L6</td>
<td>2.62</td>
<td>2.72</td>
<td>0.35</td>
<td>4.09</td>
</tr>
<tr>
<td>L7</td>
<td>2.90</td>
<td>1.29</td>
<td>2.60</td>
<td>-1.03</td>
</tr>
<tr>
<td>L8</td>
<td>-3.26</td>
<td>0.37</td>
<td>-0.52</td>
<td>-4.29</td>
</tr>
<tr>
<td>L9</td>
<td>-0.41</td>
<td>0.21</td>
<td>-0.02</td>
<td>2.80</td>
</tr>
<tr>
<td>Global</td>
<td>0.26</td>
<td>0.46</td>
<td>0.27</td>
<td>0.35</td>
</tr>
</tbody>
</table>

(K) improved some AUC parameters. However, the Bilateral (B) filter did not increase any AUC parameter.

The absolute analysis is better conducted by generating a sensitivity index. This index was calculated by dividing the rate RAD by the rate RFD. So, Table 4 was build using Tables 2 and 3.

6 CONCLUSIONS

Addressing the Initial Global Localization Problem, this paper presented a new approach to solve such problem. Another contribution from this work, is to show situations in which the pre-filtering step can reduce the computational time without reduce the characterization power from a classifier or a global localization model.

It was achieved 80% in AUC parameters from the global ROC curve without using pre-filtering. After pre-filtering the images with Wiener filter it was achieved 77% in AUC parameter from the global ROC curve and reduction of 13% in the number of originally extracted features from the places.

Comparing the four filters tested, using Table 4, we concluded that the global classification power of the model is affected almost in the same way by Wiener and Gaussian. However, Gaussian can locally increase the AUC parameters better than Wiener.

On the other hand, the global classification power of the model is largely affected by the Bilateral and Kuwahara filters. Nevertheless, in our experiments the Bilateral filter did not locally increase the characterization power of the model.

References


