Influence of image compression on object detection in natural images segmented with mean shift

ANA GAŠIć1, HRVOJE D UJMIĆ1, HRVOJE TURIĆ2, VLADAN PAPIĆ1
Department of electronics1, Department of polytechnics2
University of Split
R. Boskovica b.b.
CROATIA
ana.gasi@fesb.hr, hdujmic@fesb.hr, turich@pmfst.hr, vpapic@fesb.hr

Abstract: - One of important image processing applications is detection of humans for non-urban Search and Rescue. Authors have previously developed a method which, using mean shift segmentation, searches for objects (humans and artificial objects) in natural images. It is based on assumption that such objects have different color comparing to the rest of image. Main problem in using mean shift algorithm is long processing time. In this paper, influence of image compression ratio on mean shift processing time as well as on detection results has been investigated. It is shown that increase in compression ratio results with shorter processing time while detection of objects (Recall and Precision) is not significantly deteriorated.

Key-Words: - Image segmentation, compression, mean shift, search and rescue, object detection

1 Introduction

Surveillance of various terrains in order to find some object of interest is a task that can be associated with various civil and military activities. Typical examples include man-made object detection, natural object detection and tracing moving objects.

Although some other sensors (like Synthetic Aperture Radar and Infrared camera) are also used, for such kind of tasks aerial photos are most commonly used. Aerial or long distance photos (including satellite images) are used for road detection [1], building detection [2], airplane detection [3] and other manmade objects detection. It is also used for soil-type detection, vegetation-type detection [4] and other natural objects/phenomena detection.

Human detection for the Search and Rescue (SAR) application, i.e. finding lost, hurt or persons that are in some kind of danger is among critical applications. For this type of application, motion information obtained from the acquired sequences of images (or video) is non-relevant because the individuals searched for are mainly non-moving. It means that the focus is on image processing of static images.

In SAR missions, large areas of generally unfamiliar terrain must be thoroughly inspected. As a consequence, such missions can last for several days and can demand large and diverse task forces and technical support. Therefore, significant financial and human resources are needed. Among those resources is aerial searching.

Autonomous inspection of the area of interest that includes some kind of image processing and/or artificial intelligence can be of significant assistance. Surprisingly, there are very limited numbers of articles regarding automatic human detection from aerial images. Moreover, except searching for crashed airplanes, there are a very limited number of articles dealing with search and rescue using aerial images at all [5], [6], [7].

We have developed a method for the automatic detection of humans (and other artificial materials) in images that are taken from altitude of around 100 meters [8]. Our method is based on image segmentation and color difference between natural and non-natural materials [9]. In our previous research, various image segmentation algorithms were investigated [10] and the mean shift algorithm has been chosen. Moreover, we have developed a new two-stage approach for image segmentation that significantly speeds-up segmentation with negligible performance reduction [11].

However, further processing speed improvements should be investigated since the present system is not working in real-time. Namely, major drawback of mean shift algorithm is a high computation requirement which of course mean high processing time. The quadratic computational complexity of the algorithm is a significant problem for practical applications of this algorithm. In order to boost mean shift segmentation performance various researchers has investigated possible improvements of the basic approach [12], [13], [14].

In this paper we are using different approach. We are using basic mean shift algorithm and investigating influence of the image compression on the detection results and the speed of segmentation.
Paper is organized as follows. Section 2 presents used algorithms and methods as well as the Quality factors used for the image compression. Results are presented in section 3. Conclusions are made in section 4 and acknowledgements are given in section 5.

2 Segmentation and detection

2.1 Segmentation

Robustness and computational efficiency are critical issues in operational contexts. An algorithm must be particularly robust to scenes and conditions changes. Also, no parameterization should occur, in order be used by non-experts. Usually implementation should run in real-time and on constrained systems.

As it is previously stated, mean shift algorithm [15] has been chosen as the algorithm that demonstrated the best results regarding quality of segmentation and stability (i.e. there is no need to define number of segments). Basically, mean shift algorithm has nonparametric nature with minimal user input. It does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. Given n data points \( x_i, i = 1, ..., n \) on a \( d \)-dimensional space \( R^d \), the multivariate kernel density estimate obtained with kernel \( K(x) \) and window radius \( h \) (bandwidth) is

\[
f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left( \frac{x - x_i}{h} \right),
\]

where \( c_{k,d} \) is a normalization constant which assures \( K(x) \) integrates to 1. The modes of the density function are located at the zeros of the gradient function \( \nabla f(x) = 0 \). The gradient of the density estimator (1) is

\[
\nabla f(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x_i - x) g\left( \frac{x - x_i}{h} \right)
\]

\[
= \frac{2c_{k,d}}{nh^{d+2}} \left[ \sum_{i=1}^{n} g\left( \frac{x - x_i}{h} \right) \right] \left[ \sum_{i=1}^{n} x_i g\left( \frac{x - x_i}{h} \right) \right] - x.
\]

2.2 Object detection

Object detection is based on the evaluation of the obtained image segments. Detailed procedure and knowledge-based presuppositions that are made in order to improve correct detections and reduce false detections are presented in our previous work [8]. Also, processing speed can be improved by using divide-and-conquer approach presented in [11].

Main steps of the detection procedure are initialization, image segmentation and decision-making. More detailed description of the decision making module is presented in Fig.1.
2.3 Compression

In order to reduce size of the image, and hopefully processing time, JPEG compression has been applied on the original images. This kind of compression is lossy compression method but we hoped that because of nature of the method and our application there could be some compression ratio up to which there will be no (or negligible) degradation in overall detection results of the presented procedure.

All images were converted to standard 8-bit digitized versions of YCbCr with Cb and Cr scaled to a range of 16 to 240. JPEG compression gives priority to brightness details over fine color details and as a consequence, the resolution of the chroma components (Cb and Cr) is reduced by a factor of 2.

The image is split into blocks of 8×8 pixels, and for each block, each of the Y, Cb, and Cr data undergoes a discrete cosine transform (DCT). The amplitudes of the frequency components are quantized.

Possible problem regarding detection quality of the compressed images is in its orientation towards human vision system limitations while we are trying to make the most of the computer-based image processing advantages. Particularly, because human vision is much more sensitive to small variations in color or brightness over large areas than to the strength of high-frequency variations, the magnitudes of the high-frequency components are stored with a lower accuracy than the low-frequency components.

To change the quality (i.e. compression) of images we have used Quality factor defined by Independent JPEG Group [16] (defined on a scale of 1-99, 99 is best quality and 1 is worst).

Note that Quality factors (QF) are not percentages, nor is there a direct correlation with the final file size.

The Quality factor affects to what extent the resolution of each frequency component is reduced. Usually, high quality image can be obtained with QF 75, good quality image with QF 50, while QF under 20 results with inadequate quality (for all 3 cases, Quality factor refers to Y component).

We have used several (9) typical Quality factors that are presented in Table 1, separately for Y component and separately for Cb and Cr component. Corresponding Quality factor for Cb and Cr is higher than Quality factor for Y because Cb and Cr were already sub sampled by factor of 2.

<table>
<thead>
<tr>
<th>Y</th>
<th>97.8</th>
<th>82.9</th>
<th>67.7</th>
<th>51.5</th>
<th>38.6</th>
<th>31.2</th>
<th>26.0</th>
<th>22.2</th>
<th>19.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cb, Cr</td>
<td>98.4</td>
<td>90.6</td>
<td>82.3</td>
<td>73.4</td>
<td>64.5</td>
<td>56.2</td>
<td>47.5</td>
<td>40.6</td>
<td>35.7</td>
</tr>
</tbody>
</table>

Table 1. Quality factors for Y and Cb, Cr
time (absolute and relative) for different Quality factors is given in Table 3. Relative processing time for different Quality factors is also schematically given on Fig. 1. Of course, image coding into JPEG format takes some time. But, comparing to time needed for mean shift, this time can be neglected, especially because it can be done during image capturing on camera.

<table>
<thead>
<tr>
<th>Quality factor</th>
<th>number of objects</th>
<th>correct</th>
<th>false</th>
<th>miss</th>
<th>recall</th>
<th>precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>97.8</td>
<td>29</td>
<td>25</td>
<td>4</td>
<td>4</td>
<td>86.2</td>
<td>86.2</td>
</tr>
<tr>
<td>82.9</td>
<td>29</td>
<td>23</td>
<td>4</td>
<td>6</td>
<td>79.3</td>
<td>85.2</td>
</tr>
<tr>
<td>67.7</td>
<td>29</td>
<td>24</td>
<td>4</td>
<td>5</td>
<td>82.8</td>
<td>85.7</td>
</tr>
<tr>
<td>51.5</td>
<td>29</td>
<td>23</td>
<td>3</td>
<td>6</td>
<td>79.3</td>
<td>88.5</td>
</tr>
<tr>
<td>38.6</td>
<td>29</td>
<td>22</td>
<td>3</td>
<td>7</td>
<td>75.9</td>
<td>88.0</td>
</tr>
<tr>
<td>31.2</td>
<td>29</td>
<td>24</td>
<td>4</td>
<td>5</td>
<td>82.8</td>
<td>85.7</td>
</tr>
<tr>
<td>26.0</td>
<td>29</td>
<td>23</td>
<td>3</td>
<td>6</td>
<td>79.3</td>
<td>88.5</td>
</tr>
<tr>
<td>22.2</td>
<td>29</td>
<td>23</td>
<td>5</td>
<td>6</td>
<td>79.3</td>
<td>82.1</td>
</tr>
<tr>
<td>19.6</td>
<td>29</td>
<td>24</td>
<td>7</td>
<td>5</td>
<td>82.8</td>
<td>77.4</td>
</tr>
</tbody>
</table>

Table 2. Comparison of recall and precision for different Quality factors.

<table>
<thead>
<tr>
<th>Quality factor</th>
<th>relative processing time (%)</th>
<th>average processing time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>97.8</td>
<td>100</td>
<td>574</td>
</tr>
<tr>
<td>82.9</td>
<td>74.8</td>
<td>429</td>
</tr>
<tr>
<td>67.7</td>
<td>72.0</td>
<td>412</td>
</tr>
<tr>
<td>51.5</td>
<td>85.8</td>
<td>490</td>
</tr>
<tr>
<td>38.6</td>
<td>84.0</td>
<td>466</td>
</tr>
<tr>
<td>31.2</td>
<td>79.3</td>
<td>454</td>
</tr>
<tr>
<td>26.0</td>
<td>71.6</td>
<td>410</td>
</tr>
<tr>
<td>22.2</td>
<td>68.8</td>
<td>395</td>
</tr>
<tr>
<td>19.6</td>
<td>42.8</td>
<td>245</td>
</tr>
</tbody>
</table>

Table 3. Comparison of average processing time for different Quality factors.

There are altogether 29 objects in the 22 images including 5 images without any object. Out of 29 objects there are actually 20 humans and 9 other objects. Usually, in the SAR missions, the searching targets are humans, but sometimes looking for some other objects is useful as well (for example car, backpack, jacket, etc).

One of the original images (Quality factor 97.8) with a person, climber on the rock face, is shown on Fig.2. Enlarged part of the resulting image that contains object (with marked detected object) is shown on Fig.3 for Quality factor 97.8 and on Fig.4 for Quality factor 26.0.

Enlarged part of segmented image can be seen on Fig.5. (Quality factor 97.8) and Fig.6 (Quality factor 26.0). For this particular example, altogether 9 clusters were found for QF 97.8, and 6 for QF 26.0.

Note that, due to the nature of mean-shift algorithm, the number and composition of the clusters may be different on each execution on the same image. However, these differences occur very rarely and may be neglected especially considering particular application.
Results show that processing time drops with increasing compression rate i.e. with lowering Quality factor. For example, for image compressed with Quality factor 19.6 processing time is only 42.8 % comparing to original image processing time. It should be mentioned that original image has also been compressed but with very high Quality factor of 97.8. At the same time, Recall and Precision has not been significantly changed, meaning that detection quality hasn’t significantly deteriorated. Instead of 25 detected objects, the worst case yields 22 detected objects.

This shows that even in the case of very high compression, detection capability of small color-distinguished objects has been preserved while processing time drops significantly.

4 Conclusion

In this paper, influence of image compression rate on mean-shift segmentation performance has been investigated. The authors have previously developed a method for detection of artificial objects in natural images taken from altitudes of around 100 m. Purpose of application was human detection for non-urban Search and Rescue. In the segmentation part of the method, mean shift algorithm has been used because it showed the best detection results. However, the known problem of the mean shift algorithm is long processing time.

We showed that increasing compression of the images reduces time needed for mean shift segmentation up to 43% of original time. At the same time, Recall and Precision hasn’t significantly deteriorated.

Our research showed that even in the case of very high compression, detection capability of small color-distinguished objects has been preserved while processing time drops significantly. So, compressed images can be used for detection of non-natural objects in natural images.

5 Acknowledgement

This work was supported by the Ministry of Science and Technology of the Republic of Croatia under projects: Computer Vision in Identification of Sport Activities Kinematics (177-0232006-1662) and ICT systems and services based on integration of information (023-0231924-1661).

Special thanks to Croatian Mountain Rescue Service and Croatian Air Forces.

References:


[16] Independent JPEG Group, www.iijg.org