IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903044 |

Text Detection and Recognition from Natural Scene

S. Haritha, S Yoga.,Msc.,M.Phil

Student, Dept. of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram, TamilNadu, India

Assistant Professor, Dept. of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram, TamilNadu, India

ABSTRACT: Many techniques and algorithm have been developed to solve the problem of text extracted from natural scenes. Text extraction is emerging and challenging era in the computer vision. Text which is embedded into the image contains semantic information which is used in many other applications such as information retrieval of complex images, robot navigation, useful for visually impaired persons, street signs, automatic read the sign board and use in so many other applications. Most of the research work in this area has been done only on printed text, a very few research is addressing the LED scene text. Scene text is difficult to extract due to blur image, variations in color, noise problem, complex background, discontinuity, poor lighting conditions, and variation in illumination. LED is which is widely used in displaying the information in boards. Now days LED display that is natural scene is being widely used for displaying announcements, sign boards, banners for displaying information. To extract the text from the LED display is not an easy task, it is very complex due to its discontinuity. So, the aim of this paper to propose a technique to extract the two type of LED text from natural scene image. The first step of the algorithm is preprocessing of the image where the image is converted from RGB to grayscale, noise is removed and the image is converted to binary image, etc. Then the text is localized. After that connected component approach is used for text detection and finally the text has been recognized using template matching with correlation. The experimental results of the proposed method show the detection and recognition rate is 82.87 and 57.6.

I. INTRODUCTION

In electrical engineering and computer science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging. With the increasing use of digital image capturing devices, such as digital cameras, mobile phones and PDAs, content-based image analysis techniques are receiving intensive attention in recent years. Among all the contents in images, text information has inspired great interests, since it can be easily understood by both human and computer, and finds wide applications such as license plate reading, sign detection and translation, mobile text recognition, content-based web image search. This paper presents a real application to intelligent transportation systems (ITS) of a method to detect and recognize text in images taken from natural scenarios proposed by the same authors. This text reading algorithm has proved to be robust in many kinds of real-world scenarios, including indoor and outdoor places with a wide variety of text appearance due to different writing styles, fonts, colors, sizes, textures and layouts, as well as the presence of geometrical distortions, partial occlusions, and different shooting angles that may cause deformed text.

In this paper, this algorithm is applied, including some modifications and new functionalities, to read the information contained in traffic panels using the images served by Google Street View. The aim of this work is, in the first place, to detect traffic panels and to recognize the information inside them, showing that the text detection and recognition method proposed in [1] can be generalized to other scenarios, which are completely different to those that have been tested, without needing to retrain the system. In the second place, we want to develop an application that enables the creation of up-to-date inventories of traffic panels of regions or countries that facilitate traffic signposting maintenance and driver assistance. In this paper, we focus on traffic panels in the Spanish territory for two main reasons. First, unlike other countries, the coverage of Street View in Spain is near complete; thus, we can create a huge and diverse data set of images. Second, as far as we know, there is not any official database of all the traffic panels in Spain; thus,

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903044 |

there are more possibilities that any government or institution responsible for managing the road network can be interested in having an up-to-date inventory of the traffic panels in Spain with the method here proposed.

The reasons for which these organizations may be interested are various. Having a centralized database of all the traffic panels supposes a rapid and economic way of evaluating and analyzing the potential dangerous situations that may arise due to traffic panels that suffer from a bad visibility or show deteriorated or outdated information. Street-level panoramic image recording services, such as Street View, which have become very popular in the recent years and have reached a huge coverage of the road network, suppose a potential source to rapidly know the state of the vertical signposting of the road network, particularly when the street-level images are updated regularly. Computer vision techniques applied on this kind of images simplify and speed up the creation of traffic signposting inventories, minimizing the human interaction. In addition, these inventories can be useful not only for supporting maintenance but also for developing future driver assistance systems. In general, automatic text read-ing may be helpful to support drivers or autonomous vehicles to find a certain place by simply reading and interpreting street signs, road panels, variable-message signs, or any kind of text present in the scenario, when global positioning systems suffer from lack of coverage, particularly in high-density urban areas.

II. LITERATURE SURVEY

Fast Approximate Energy Minimization via Graph Cuts

In this paper they address the problem of minimizing a large class of energy functions that occur in early vision. The major restriction is that the energy function's smoothness term must only involve pairs of pixels. They propose two algorithms that use graph cuts to compute a local minimum even when very large moves are allowed. The first move they consider is a swap: for a pair of labels, this move exchanges the labels between an arbitrary set of pixels labeled and another arbitrary set labeled. Their first algorithm generates a labeling such that there is no swap move that decreases the energy. The second move they consider is an -expansion: for a label, this move assigns an arbitrary set of pixels the label. Their second algorithm, which requires the smoothness term to be a metric, generates a labeling such that there is no expansion move that decreases the energy.

Moreover, this solution is within a known factor of the global minimum. They experimentally demonstrate the effectiveness of their approach on image restoration, stereo and motion. The major difficulty with energy minimization for early vision lies in the enormous computational costs. Typically these energy functions have many local minima (i.e., they are non-convex). Worse still, the space of possible labeling has dimension jPj, which is many thousands. There have been numerous attempts to design fast algorithms for energy minimization. Simulated annealing was popularized in computer vision, and is widely used since it can optimize an arbitrary energy function.

Unfortunately, minimizing an arbitrary energy function requires exponential time, and as a consequence simulated annealing is very slow. In practice, annealing is inefficient partly because at each step it changes the value of a single pixel. The energy functions that they consider in this paper arise in a variety of different contexts, including the Bayesian labeling of MRF's. The algorithms described in this paper generalize the approach that they originally developed for the case of the Potts model.

In particular, they compute a labeling which is a local minimum even when very large moves are allowed. They begin with an overview of their energy minimization algorithms, which are based on graph cuts. Their first algorithm, is based on -swap moves and works for any semi metric Vfp;qg's. Their second algorithm, described, is based on more interesting –expansion moves but works only for metric Vfp;qg's (i.e., the additional triangle inequality constraint is required). Note that -expansion moves produce a solution within a known factor of the global minimum of E.

III. SYSTEM ANALYSIS

3.1 Existing System

In existing system present a hybrid approach to robustly detect and localize texts in natural scene images by taking advantages of both region-based and CC-based methods. Since local region detection can robustly detect scene texts even in noisy images, they design a text region detector to estimate the probabilities of text position and scale. Although the existing methods have reported promising localization performance, there still remain several problems to solve. For region-based methods, the speed is relatively slow and the performance is sensitive to text alignment orientation. On the other hand, CC-based methods cannot segment text components accurately without prior knowledge

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903044 |

of text position and scale. Here, designing fast and reliable connected component analyzer is difficult since there are many non-text components which are easily confused with texts when analyzed individually.

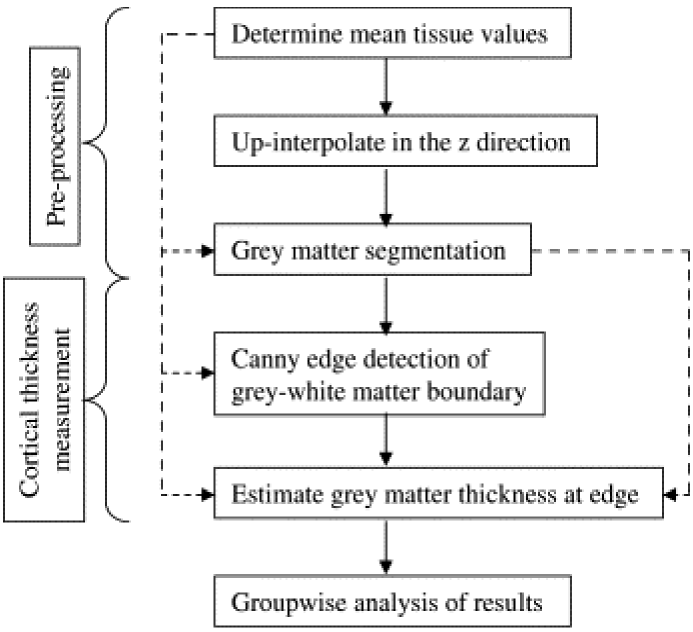
3.2 Proposed System

In our proposed system has a selective metric-based clustering to extract textual information in real-world images. Hence, our selective metric-based clustering is integrated into a dynamic method suitable for text extraction and character segmentation. This method uses several metrics to merge similar color together for an efficient text-driven segmentation in the RGB color space. However, color information by itself is not sufficient to solve all natural scene issues; hence we complement it with intensity and spatial information obtained using Log–Gabor filters, thus enabling the processing of character segmentation into ndividual components to increase final recognition rates. Our selective metric-based clustering uses mainly color information for text extraction and our system fails for natural scene images having embossed characters. In this case, foreground and background have the same color presenting partial shadows around characters due to the relief but not enough to separate textual foreground from background in a discriminative way as displayed. Gray-level information with the simultaneous use of a priori information on characters could be a solution to handle these cases. Next we propose a new text validation measure M to find the most textual foreground cluster over the two remaining clusters. Based on properties of connected components of each cluster, spatial information is already added at this point to find the main textual cluster. The proposed validation measure, M, is based on the largest regularity of connected components of text compared to those of noise and background. And also we use Log–Gabor filters that present globally high responses to characters. Hence, in order to choose efficiently which clustering distance is better to handle text extraction, we perform an average of pixel values inside each mask. The mask which has the highest average is chosen as the final segmentation.

IV. SYSTEM IMPLEMENTATION

4.1 Implementation of the pre-processing stage

In this module we implement the preprocessing stage of the overall process. At the preprocessing stage, a text region detector is designed to detect text regions in each layer of the image pyramid and project the text confidence and scale information back to the original image, scale-adaptive local binarization is then applied to generate candidate text components. To extract and utilize local text region information, a text region detector is designed to estimate the text confidence and the corresponding scale, based on which candidate text components can be segmented and analyzed accurately. Initially, the original color image is converted into a gray level image. To measure the text confidence for each image patch in a window, no matter it is accepted or rejected. The text scale map is used in local binarization for adaptively segmenting candidate CCs and the confidence map is used later in CCA for component classification. They calculate the radius from the text scale map which is more stable under noisy conditions. After local binarization, because we assume that within each local region, gray-level values of foreground pixels are higher or lower than the average intensity.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

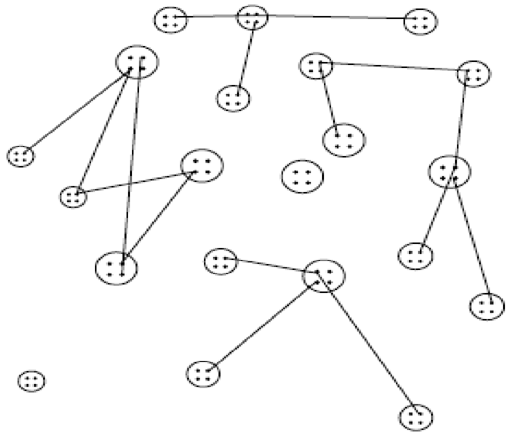
|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903044 |

A preprocessor is a program that processes its input data to produce output that is used as input to another program. The output is said to be a preprocessed form of the input data, which is often used by some subsequent programs like compilers. The amount and kind of processing done depends on the nature of the preprocessor; some preprocessors are only capable of performing relatively simple textual substitutions and macro expansions, while others have the power of full-fledged languages.

4.2 Implementation of the connected component analysis stage

In this module we implements the connected component analysis (CCA) stage, a CRF model combining unary component properties and binary contextual component relationships is used to filter out non-text components. Here we propose a conditional random field (CRF) model to assign candidate components as one of the two classes ("text" and "non-text") by considering both unary component properties and binary contextual component relationships. CRF is a probabilistic graphical model which has been widely used in many areas such as natural language processing. Next considering that neighboring text components normally have similar width or height, we build up a component neighborhood graph by defining a component linkage rule. And also we use the CRF model to explore contextual component relationships as well as unary component properties. During the test process, to alleviate the computation overhead of graph inference, some apparent non-text components are first removed by using thresholds on unary component features. The thresholds are set to safely accept almost all text components in the training set.



Connected-component labeling is not to be confused with segmentation. Connected-component labeling(stage) is used in computer vision to detect connected regions in binary digital images, although color images and data with higher dimensionality can also be processed. When integrated into an image recognition system or human-computer interaction interface, connected component labeling can operate on a variety of information. Blob extraction is generally performed on the resulting binary image from a thresholding step. Blobs may be counted, filtered, and tracked.

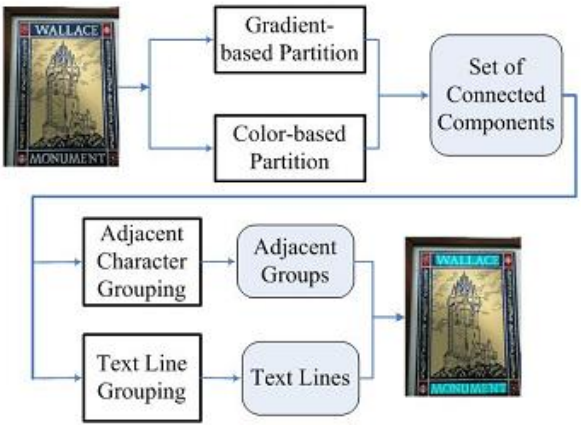
4.3 Implementation of text grouping method

In this module we implements text grouping method. To group text components into text regions are lines and words, we design a learning-based method by clustering neighboring components into a tree with a minimum spanning tree (MST) algorithm and cutting off between-line (word) edges with an energy minimization model. we cluster text components into a tree with MST based on a learned distance metric, which is defined between two components as a linear combination of some features. With the initial component tree built with the MST algorithm, between-line/word edges need to be cut to partition the tree into subtrees, each of which corresponds to a text unit. Finally, text words corresponding to partitioned subtrees can be extracted and the ones containing too small components are removed as noises.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903044 |



4.4 Implementation of selective metric-based clustering using Log–Gabor filters

In this module we implement the selective metric based clustering using log-Gabor filter. Hence, our selective metric-based clustering is integrated into a dynamic method suitable for text extraction and character segmentation. This method uses several metrics to merge similar color together for an efficient text-driven segmentation in the RGB color space. However, color information by itself is not sufficient to solve all natural scene issues; hence we complement it with intensity and spatial information obtained using Log–Gabor filters, thus enabling the processing of character segmentation into individual components to increase final recognition rates. Our selective metric-based clustering uses mainly color information for text extraction and our system fails for natural scene images having embossed characters. In this case, foreground and background have the same color presenting partial shadows around characters due to the relief but not enough to separate textual foreground from background in a discriminative way as displayed. Gray-level information with the simultaneous use of a priori information on characters could be a solution to handle these cases. Next we propose a new text validation measure M to find the most textual foreground cluster over the two remaining clusters. Based on properties of connected components of each cluster, spatial information is already added at this point to find the main textual cluster. The proposed validation measure, M, is based on the largest regularity of connected components of text compared to those of noise and background. And also we use Log–Gabor filters that present globally high responses to characters. Hence, in order to choose efficiently which clustering distance is better to handle text extraction, we perform an average of pixel values inside each mask. The mask which has the highest average is chosen as the final segmentation.

log-Gabor wavelets increases with the number of orientations (real parts in the left column and imaginary parts in the right column).in which (ρ,θ) are the log-polar coordinates (octave scales), k indexes the scale and p is the orientation, the pair $(\rho_k; \theta_{pk})$ corresponds to the frequency center of the filters, and $(\sigma\rho; \sigma\theta)$ the angular and radial bandwidths.

$$G_{pk} = G(\rho,\theta,p,k) = \exp\left(-\frac{1}{2}\left(\frac{\rho-\rho_k}{\sigma_\rho}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{\theta-\theta_{pk}}{\sigma_\theta}\right)^2\right)$$

$$\text{with}\quad \begin{cases} \rho_k = \log_2(n) - k \\ \theta_{pk} = \begin{cases} \frac{\pi}{P}p & \text{if } k \text{ is odd} \\ \frac{\pi}{P}(p+\frac{1}{2}) & \text{if } k \text{ is even} \end{cases} \\ (\sigma_\rho,\sigma_\theta) = 0.996(\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{2}}\frac{\pi}{P}) \end{cases}$$

The main particularity of this novel scheme is the construction of the low-pass and high-pass filters. Such a complete scheme approximates flat frequency response and therefore exact image reconstruction which is obviously beneficial for applications in which inverse transform is demanded, such as texture synthesis, image restoration, image fusion or image compression.

4.5 Performance evaluation

Finally in this module the proposed approaches were illustrated and evaluated to compare the performance of all the approaches. We analyze our proposed scheme in terms of extraction rate, precision rate recall rate and average

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

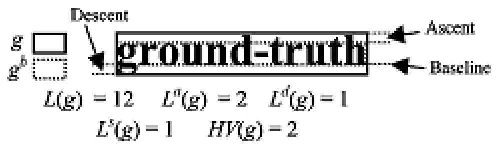
|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903044 |

speed. Based on the comparison and the results from the experiment show the proposed approach works better than the other existing systems. performance evaluation based on following steps.

1. **Textbox Location:**

$$(\mathrm{Left}(g),\ \mathrm{Top}(g),\ \mathrm{Right}(g)\ \text{and}\ \mathrm{Bottom}(g)).$$



2. **Textbox Height:**

$$\text{Equal to } (\mathrm{Bottom}(g) - \mathrm{Top}(g) + 1).$$

3. **Text String:**
   The ASCII string of the text in g(it is not used in our protocol)
4. **Text Length :**
   a. The number of the characters (i.e., the length of the text string) in g .
5. **Skew Angle :**
   a. The skew angle of the text font in . Typically, the greater the skew angle is, the more difficult the textboxes that can be detected.
6. **Color and Texture :**
   a. If the text string in is nonhomochromous or textured, is set to 1; otherwise, it is set to zero. Generally, homochromous text strings are easier to correctly detect.
7. **Background Complexity :**
   a. We extend in each direction for ten pixels and thus form an extended textbox denoted by g^e . The background complexity of is g described by

V. CONCLUSION AND FUTURE ENHANCEMENT

5.1 CONCLUSION

This project focuses on implement the selective metric-based clustering uses mainly color information for text extraction and our system fails for natural scene images having embossed characters. In this case, foreground and background have the same color presenting partial shadows around characters due to the relief but not enough to separate textual foreground from background in a discriminative way as displayed. And also we use Log–Gabor filters that present globally high responses to characters. Hence, in order to choose efficiently which clustering distance is better to handle text extraction, we perform an average of pixel values inside each mask. Finally the proposed approaches were illustrated and evaluated to compare the performance of all the approaches.

5.2 FUTURE ENHANCEMENT

In this project we detect the text region using selective metric based clustering, which provides high probability of text detection. Also we used connected component analysis and text grouping and relates the probability of output and came out with an high possibility of text detection. As an future enhancement we are going to implement text extraction and achieve it for mobile application.

REFERENCES

[1] C. M. Bishop, Neural Networks for Pattern Recognition. New York: Oxford University Press, 1995.
[2] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 11, pp. 1222–1239, 2001.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903044 |

[3] D. T. Chen, J.-M. Odobez, and H. Bourlard, "Text detection and recognition in images and video frames," Pattern Recogn., vol. 37, no. 5, pp. 595–608, 2004.
[4] X. L. Chen, J. Yang, J. Zhang, and A.Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.
[5] X. R. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'04), Washington, DC, 2004, pp. 366–373.
[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, 2005, pp. 886–893.
[7] S. L. Feng, R. Manmatha, and A. McCallum, "Exploring the use of conditional random field models and HMMs for historical handwritten document recognition," in Proc. 2nd Int. Conf. Document Image Analysis for Libraries (DIAL'06), Washington, DC, 2006, pp. 30–37.
[8] J. Gllavata, R. Ewerth, and B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients," in Proc. 17th Int. Conf. Pattern Recognition (ICPR'04), Cambridge, U.K., 2004, pp. 425–428.
[9] J. M. Hammersley and P. Clifford, Markov Field on Finite Graphs and Lattices, 1971, unpublished.
[10] X.-B. Jin, C.-L. Liu, and X. Hou, "Regularized margin-based conditional log-likelihood loss for prototype learning," Pattern Recogn., vol. 43, no. 7, pp. 2428–2438, 2010.
[11] Y. Jin and S. Geman, "Context and hierarchy in a probabilistic image model," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'06), New York, NY, 2006, pp. 2145–2152.
[12] B.-H. Juang and S. Katagiri, "Discriminative learning for minimum error classification," IEEE Trans. Signal Process., vol. 40, pp. 3043–3054, 1992.
[13] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: A survey," Pattern Recogn., vol. 37, no. 5, pp. 977–997, 2004.
[14] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, 2003.
[15] S. Kumar, J. August, and M. Hebert, "Exploiting inference for approximate parameter learning in discriminative fields: An empirical study," in Proc. 5th Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern (EMMCVPR'05), St. Augustine, FL, 2005, pp. 153–168.