Proceedings of the 3rd International Conference on Vision Systems, Graz, Austria, April 1-2, 2003, Springer-Verlag, Berlin, LNCS 2626, pp. 141-150

Alignment Of Sewerage Inspection Videos for Their Easier Indexing

Karel Hanton, Vladimír Smutný, Vojtěch Franc, and Václav Hlaváč

Czech Technical University, Faculty of Electrical Engineering
Department of Cybernetics, Center for Machine Perception
121 35 Prague 2, Karlovo náměstí 13, Czech Republic
{hanton,smutny,xfrancv,hlavac}@cmp.felk.cvut.cz
http://cmp.felk.cvut.cz

Abstract. The paper describes a new module of the developed robotic sewerage inspection system. The sewerage pipe is inspected by a remotely controlled inspection tractor equipped by a camera head able to rotate and zoom. This contribution describes a method and a software solution which allows to align the new inspection video and the archived video of the same pipe section (typically captured ten years ago). The aim of the analysis is to see how the pipe defects develop in time.
The alignment of videos based on correspondences sought in images is overambitious. We have chosen the pragmatic approach. The text information from odometer which is superimposed in the video is automatically located and recognized using Optical Character Recognition (OCR) technique. The recognized distance from man-hole of the pipe allows to align both videos easily. The sewerage rehabilitation expert can then use only one remote control of the VCR for video positioning.
This contribution describes the proposed solution, briefly mentions its implementation and demonstrate its function on practical sewerage inspection videos. However, our indexing approach can be used with any videos with superimposed text.

1 Introduction and problem formulation

This paper reports about the subproblem[1] studied within the EU Take-up project ISAAC (Inspecting Sewerage Systems and Image Analysis by Computer), IST-2001-33266, running from January till December 2002. The aim of the ISAAC project is to transfer several computer vision technologies to an established robotics area – sewerage inspection by the remotely operated inspection tractor equipped with a TV camera. The sewerage bylaws in some countries require that each sewerage section is inspected every ten years at least.

[1] Acknowledgement: This research was supported by the EU project ISAAC, IST-2001-33266, Czech Ministry of Education, project LN00B096. The part related to the development of multiclass SVM classifier was supported by the EU project ActIPret, IST-IST-2001-32184 and CTU grant 0208313.

The tractor operator captures a VHS video of the examined sewerage section. He/she seeks for defects, e.g. cracks in the sewerage walls. There is an important step which compares the newly captured inspection video with the video from the same sewerage section captured several years ago. This is performed off-line in the laboratory. The most important result of the analysis is the assessment how sewerage defects develop in years. This information is essential to managers who prepare the plan of sewerage repairs.

Our end-user in the ISAAC project, the Prague Water and Sewerage Company, also has to compare new and old videos. Their inspection tractor augments text information to the video stream. One of it is the position of the robot in the pipe obtained from the inspection tractor. There is an odometer which measures the distance information from the cable which is pulled by the tractor. The distance from odometer is displayed as a text in the video.

When the operator wants to compare old and new video about the same sewerage section he would use two VCRs. It is likely that the two videos were taken by two different equipments and several years apart. The were likely taken with different zoom and direction of view. The result is that it is almost impossible to use only the image content for video alignment. The operator aligns videos by operating two remote controllers simultaneously. The operator visually compares two images and evaluates defects and their development in time. This work is tedious and error-prone as operators spend most of their time and efforts by watching uninteresting parts of videos.

This paper describes a simple but powerful trick. The text information from the odometer, which is superimposed in the video, is automatically located in image frames and recognized using OCR techniques. The recognized distance from man-hole of the pipe allows easily to align both videos. The operator can then use only one remote control for video positioning.

The paper is organized as follows. Section 2 informs about the state-of-the-art. Section 3 describes the proposed method. Section 4 reports about implementation and experimental results. Section 5 draws conclusions.

2 State-of-the-art

To our knowledge the visual comparison of old and new sewerage inspection analog videos is performed manually as described above in most of the waste water treatment organizations.

Digital videos have been becoming popular in recent years. The digital video can store additional information including distance information from the odometer. It is likely that in the future the distance information will be recorded and will allow an easy indexing. However, due to the need to compare with old analog videos it is foreseen that the proposed method will be needed in the future too.

3 Proposed solution

Our problem of comparing two video sequences (tapes) can be transformed to a more general problem of locating desired frame in the video sequence. This desired image frame is specified by some indexing parameters.

This case embodies our 'comparison' problem because the parameters can be acquired from the second (archive) video sequence. There are several such indexing parameters as, e.g., the frame number, the time mark, auxiliary information placed in the frame image, etc.

In our case of comparing two sequences of sewerage inspections, the natural parameter indexing the video sequence is the position of the inspecting tractor in the pipe. The position is meant as an absolute distance along the pipe from the man-hole (coordinate origin) through which the robot was inserted to the pipe. Having such location of the inspection tractor, the practical issues can be addressed, e.g.: What is the distance to the lateral pipe from the man-hole? Move the tractor to the location X.

The problem was transformed from comparing two video sequences to determining the distance (a single parameter) between the actual robot location in which the corresponding image frame was captured and a beginning of the inspection tractor trajectory (most often in the man-hole). The distance can be neither derived from the frame number nor from the time mark because the two videos were likely acquired in different speed or unknown pauses were inserted. Thus the distance has to be determined directly from the content of image frames. The first natural idea is to try to 'understand' the seen pipe surface via, e.g., some markers on it as the pipe joints, laterals, defects. The use of pipe invariants (e.g., its diameter, length of segments) could be used to overcome the problem that the images could be taken by different inspection tractors, cameras could look to different direction, have different zoom, etc. This was an overambitious idea for the ISAAC project.

The second idea was rather pragmatic. Most inspection tractors count the distance from the beginning of inspection path by odometer on the cable. This is also the case with several different tractors manufactured by the German company Rausch which have been used by the Prague Water and Sewerage Company. The distance along the inspection path is also measured and displayed by tractors manufactured by the the ISAAC project partner – the Pearpoint Ltd from the U.K. The only place where the tractor position information has been stored on analog video tapes is its superposition as the text to video frames, see Fig. 1.

The robot localization problem was converted to reading the odometry information from image frames. The good news is that for a particular inspection system the odometry information is written in the fixed position of the frame, in a priori known format, using a nonproportional font. The bad news are that quality and resolution of images on VHS-tapes is low and different systems have different form of overlaid text. However, the latter problem can be overcome by training the system for a new device.

Fig. 1. Example of the image frame from the inspection video with the text information superimposed.

The proposed text recognition consists of four steps: (1) localization of odometry information in the image frame by circumscribing it by rectangular region of interest, (2) splitting the text in the region of interest into individual characters, (3) recognition of each character, and (4) interpretation of recognized characters.

For a video from a particular inspection tractor the odometry text is on the same position. The localization is given by a rectangular window (region of interest) which is slightly bigger than the odometry information on the screen. The position and size of the window was either stored before for a particular video sequence or the user performs additional learning step, i.e. the user draws manually a rectangle encompassing the text information on her/his computer screen and specifies the format in the text. The described procedure constitutes the step (1).

In the step (2), the text in the region of interest is separated into individual characters. The characters do not overlap, are written in the nonproportional font in the fixed raster. The problem is undersampling as the width of the raster cell for one character has been typically from 10 to 14 pixels in our experiments. The quality of the image digitized from the analog video tape is low too. The method used to segment individual characters is described in Section 3.1.

Step (3) recognizes individual characters into twelve classes using Support Vector Machine classifier. This is performed by the SVM classifier, which classifies each character to one of twelve classes: numerals 0-9, full stop, and character m (for meters). The classification is explained in Section 3.2.

Step (4) interprets the recognized text. This allows to perform the consistency check. The text has a fixed syntax. For example for the German sewerage inspection system Rausch we expect number in followed format [*?.??m] (in regular expression notation). There are additional constraints stemming from the

context given by neighboring video frames. For instance, due to certain maximal speed of the inspection tractor the difference between two consequent image can be a few centimeters at most.

Let us describe steps (2) and (3) in more detail in the following two sections.

3.1 Segmenting text of interest into individual characters

Step (1) provided the rectangular region of interest encompassing the text. Step (2), which is being described now, has to segment individual characters. The character separation method depends on the way the odometry information is superimposed to the video. The all different inspection tractors we have seen so far superimpose white characters on the mostly darker video background.

The rectangle encompassing text is resized to fit most tightly around the text. The thresholding with automatically set threshold from the histogram was used to separate the text from the background. The minimal bounding box is calculated. The obtained rectangle is enlarged in horizontal direction both to the left and to the right by 2 pixels to be sure that the text does not start, resp. finishes in the first, resp. last, column of the rectangle. This extended rectangle constitutes the new region of interest. The underlying grey scale image G constitutes the input information for the OCR procedure. As the inspection tractors use color cameras the intensity has to be calculated from RGB values, $G = 0.29I_R + 0.58I_G + 0.11I_B$.

The intensity profile obtained by summing up (squared) intensities along the columns is used to uncover the regular grid underlying the text. The periodicity of the profile can be unveiled.

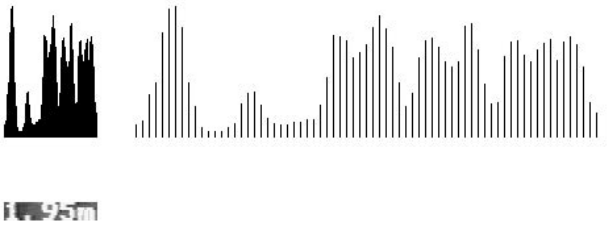


Fig. 2. Example of analyzed text and intensity profile. The original image G with the text is bottom left. Above it (top left) is the profile H. The same profile stretched so that the periodicity is visible better is top right..

The following intensity profile H, see Fig. 2, is calculated for the image $G(i, j)$ with characters, where i is the row index and j is the column index of an individual pixel. For each column j, the value of the profile is given by the sum

of squared intensities,

$$H_j = \sum_k G^2(k, j) \, .$$

The intensity values were squared when contributing to the profile because it better indicates which column is occupied by the character (higher intensity value) and which column corresponds to the inter-character space (lower intensity value). For example, if there are two neighboring zeros in the image then the space between these zeros is lighter than the background in general. This is caused by the influence of neighboring pixels. If just intensities were summed along the column it could easily happen that the sum for the space was the same as the sum over the central column of the number zero because there might be only two white pixel there. Squaring the intensity value solves the problem in our experience.

The profile H shows the amount of white pixels in each column. It can be treated as a probability distribution of the event that the column belongs to the character. The desire is to unveil the width w of the character grid and the start column ϕ of the grid with respect to the region of interest given by image G. The base frequency unveiled by the one-dimensional Fourier transformation provides the width w (length of the period corresponding to the base frequency) and start column ϕ (phase).

The static component of the profile H is suppressed by subtracting the mean value μ of the profile H. The resulting new profile h is given by the vector

$$h = H - \mu(H) \, .$$

The best periodic signal fitting to the profile h is found by varying the width of the characters w in the range expected in the image, i.e. from 9 to 15 pixels in our experiments. The periodicity of character grid is found by maximizing

$$\max_w \left\{ \left(\sum_i h_i \, \sin\left(\frac{2\pi i}{w}\right) \right)^2 + \left(\sum_i h_i \, \cos\left(\frac{2\pi i}{w}\right) \right)^2 \right\} \, .$$

The value of w maximizing the above expression corresponds to the width of one cell in the character raster. Let us denote it r. The angle between the sin and cos components of the first harmonic gives the phase ϕ,

$$\phi = \arctan \left(\frac{\sum_i h_i \, sin\left(\frac{2\pi i}{r}\right)}{\sum_i h_i \, cos\left(\frac{2\pi i}{r}\right)} \right) \, .$$

The phase ϕ corresponds to the shift of the beginning of the character grid with respect to the left column of the image G, which is constitutes the region of interest. The profile h was calculated from it. The phase ϕ is proportional to the shift of the character grid from the left column of G in pixels s,

$$s = \phi \frac{r}{2\pi} \, .$$

Having this information, the image G is segmented into individual characters.

Next, the intention is to represent characters in the form which is invariant to slight geometric and radiometric changes. Such a representation decreases complexity of involved classifier learning. Individual letters are geometrically transformed into the 10×10 pixel grid using the nearest neighbor interpolation method. The underlying grey scale image is radiometrically normalized by histogram equalization.

3.2 Recognition of individual characters by the Support Vector Machine

We selected the Support Vector Machines (SVM) [6] approach for recognition task. The SVM have been shown to perform well on variety of problems such as handwritten character recognition [4]. The standard SVM are designed for dichotomic classification problem.

We used our method [2] based on the transformation of a slightly modified multi-class criterion to the single-class SVM problem, see 3(a). The single-class SVM problem is simple to optimize and a variety of sophisticated optimization algorithms can be employed. We used a modification of the Sequential Minimal Optimizer (SMO) [5] which is simple and fast. The characters to be classified are

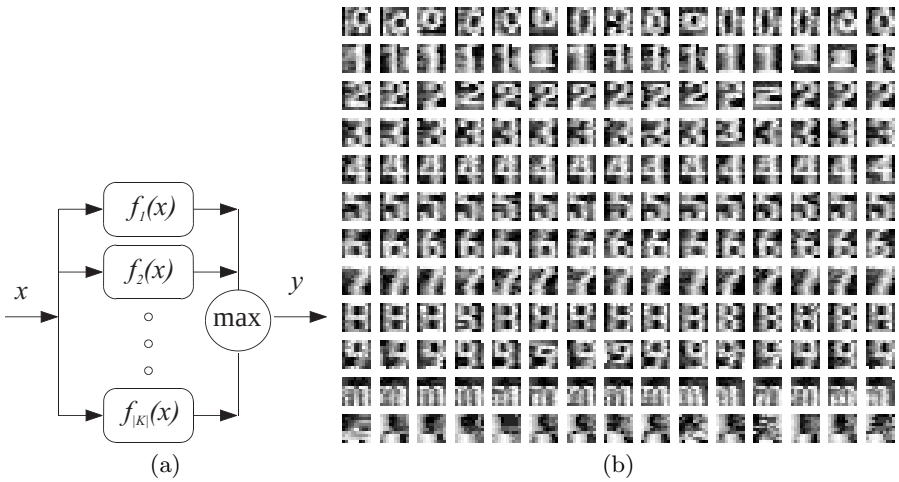


Fig. 3. (a) Classification strategy. (b) A sample of gray-scale 10×10 images of digits captured from video records.

digits $1, 2, \ldots, 9$ and characters "." and "m" presented as a 10×10 gray-scale images obtained by the procedure described above. A sample of images can be seen in Figure 3(b). The training set used contained 4331 examples of characters. The pixel gray-scale values from the interval $[0, 255]$ were mapped to the interval

$[0, 1]$. Each image is represented as a $10 \times 10 = 100$-dimensional feature vector containing intensities of corresponding pixel.

We selected Radial Basis Function (RBF) kernel $k(x, x') = e^{-0.5||x-x'||^2/\sigma^2}$ for which the training data are separable. This implies that the regularization constant can be set to infinity $C = \infty$. The only free parameter to be determined is the width σ of the RBF kernel. We computed five-fold cross-validation error rate for parameters $\sigma = [0.2, 0.5, 1, 2, 3, \ldots, 10]$. The best cross-validation 0.74% (percentage of missclassifications) was obtained for the kernel width $\sigma = 2$. The number of unique support vectors is 757 (17%). The classifier is able to process about 550 characters per second on K7/1800MHz computer. Further speed up can be achieved using approximated classification rule with less number of Support Vectors [1, 7].

4 Implementation and experiments

We implemented an experimental version of the methods that demonstrate feasibility of the approach. We currently negotiate with the customer – Prague Water and Sewerage Company about the final implementation.

The main part of the code is in the Visual C++ 6.0 under the Microsoft Windows. Video for Windows functions (AVIFile) included in Visual C++ 6.0 libraries were used to process videos. The experimental implementation has an user interface which is easy to use.

The learning part of classifier is implemented in Matlab. This implementation is part of our Statistical Pattern Recognition Toolbox which is freely available [3].

The SVM classifier parameters calculated by the Matlab code are stored to a parameter file. The main program sets up the classifier from this parameter file which defines a specific classifier tuned for a particular type of characters (font). If there is a need to classify another type of characters originating in different sewerage inspection tractor then the parameter file is replaced only.

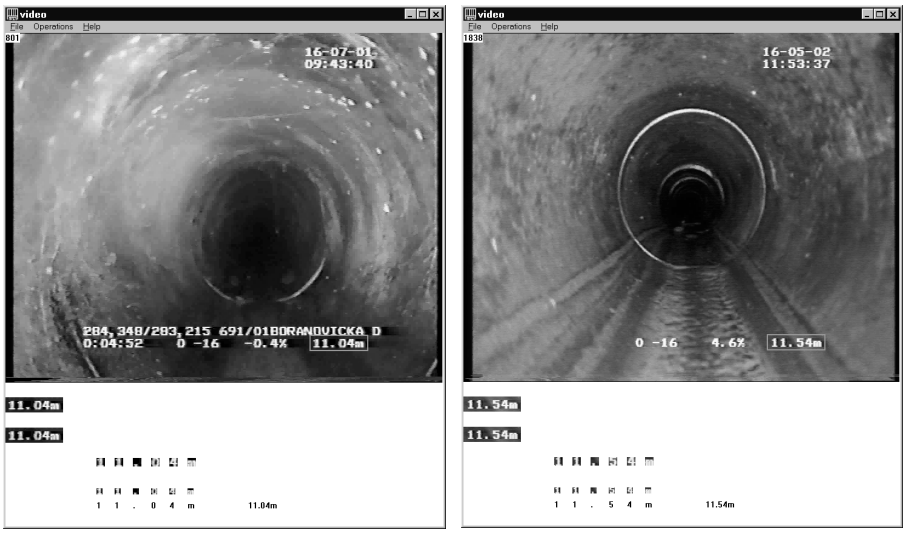
In the error correcting part, we have only checked the specific format in "*?.??m".

We performed tests on several videos from the Prague Water and Sewerage Company. Videos were captured by two two types of robots, see Fig. 4 and Fig. 5. The first robot superimposes white text on the sewerage image. The second robot uses grey rectangular area underlying the text. We have tested bad quality video from a VHS tape and the best available quality captured as a direct copy on the Super-VHS tape.

The recognition success rate even for the VHS quality was over 99%. This quality could be further improved if the error correction based on the context were implemented.

5 Conclusions and outlook

This contribution described the method which allows to align two analog videos stored on the tape provided the textual information is superimposed in image



(a) Robot 1, standard VHS. (b) Robot 1, Super VHS.

Fig. 4. (a) Screen shots showing the quality of input images and demonstrating processing.

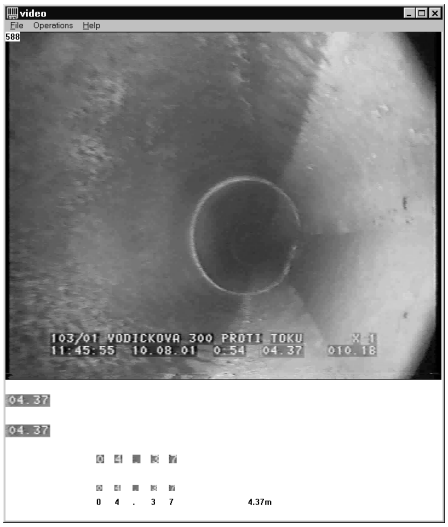


Fig. 5. Screen shot. Robot 2, 'satellite' camera on a string able to inspect lateral pipes, standard VHS showing the quality of input images and demonstrating processing.

frames. The text information is localized in the image and recognized. The recognized data serve as indexes to a video.

In our particular case, we experimented with videos captured by a camera placed in a special robotic vehicle for sewerage surveys. Such equipment is manufactured by several producers worldwide.

We described our prototype implementation which proved that the approach is feasible. The accuracy, speed and ease of handling are sufficient. The actual fully operational implementation is under negotiations.

The proposed method has more applications. The approach can be also used for converting older analog video records (e.g., VHS) to a new digital records with the indexing and content information written to additional tracks of the video record.

The approach can be used to generate automatically the interpretation database of the video provided the relevant information was superimposed as text on the video. In our particular sewerage domain it can be declination/inclination, position of lateral pipes, man-holes, images for a particular location (e.g., with the defect), etc.

References

1. C.J.C. Burges and B. Schölkopf. Improving the accuracy and speed of support vector machines. In Michael C. Mozer, Michael I. Jordan, and Thomas Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9, page 375. The MIT Press, 1997.
2. V. Franc and V. Hlaváč. Multi-class Support Vector Machine. In R. Kasturi, D. Laurendeau, and Suen C., editors, *16th International Conference on Pattern Recognition*, volume 2, pages 236–239, Los Alamitos, CA 90720-1314, August 2002. IEEE Computer Society.
3. V. Franc and V. Hlaváč. Statistical pattern recognition toolbox for Matlab, 2000-2002. http://cmp.felk.cvut.cz.
4. Y. LeCun, L. Botou, L. Jackel, H. Drucker, C. Cortes, J. Denker, I. Guyon, U. Muller, E. Sackinger, P. Simard, and V. Vapnik. Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural Networks: The Statistical Mechanics Perspective*, pages 261–276, 1995.
5. J.C. Platt. Fast training of support vectors machines using sequential minimal optimization. In B. Scholkopf, C.J.C. Burges, and A.J. Smola, editors, *Advances in Kernel Methods*. MIT Press, Cambridge, MA., USA, 1998.
6. V. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, 1998.
7. X. Xiao, H. Ai., and G. Xu. Pair-wise Sequential Reduced Set for Optimization of Support Vector Machines. In *16th International Conference on Pattern Recognition*, 2002.