

Vision and Cognitive Systems Final Project

Fabio Polito
230635@studenti.unimore.it

Giordano Costi
226934@studenti.unimore.it

Stefano Carretti
227250@studenti.unimore.it

Univerisity of Modena and Reggio Emilia

I. INTRODUCTION

In this paper, we present a method to detect and identify paintings starting from a video taken inside “Galleria Estensi, Modena”.

Each frame is processed with image processing techniques in order to localize paintings, rectify distortions, and fetch from the database the corresponding work of art.

At the same time, an artificial neural network (YoloV3) detects people, which are localized inside a room of the museum.

II. PAINTING DETECTION

All frames are extracted in sequence from the video and processed independently. The followed pipeline is explained afterwards.

A. Preprocessing

First of all, each frame is converted into black and white and processed with a bilateral filter to remove noise while preserving at the same time the edges.

After that, we apply otsu threshold with the aim to separate the pixels into background and paintings, due to their chromatic difference.

B. Bounding Box detection

Using the function `findContours` of OpenCV, we obtain the outlines of the objects in the foreground, among which there will also be the paintings that we are looking for.

We then create the bounding box containing this contours and to eliminate the rectangles identified inside the paintings, we keep only the external ones and eliminate those contained within others.

Among the remaining bounding boxes, those that are not judged as paintings by an SVM model, will get discarded.

C. SVM

To classify the ROI proposed by the previous pipeline, an SVM model is been trained.

The algorithm takes as input the concatenation of the three histograms, one for each color plane.

The training dataset is composed of 1025 instances taken from bounding boxes derivate from the videos inside the museum and manually labeled. 409 of the samples are labeled as holding a painting and 604 as inaccurate bounding box.

A radial basis function kernel is been exploited for the classification. The model returns False if the rectangle doesn't contain a painting and True if it does.

D. Precision boosting and paintings segmentation

To obtain a better segmentation within each bounding box we apply a further refinement of the images.

Observing the low light in the videos that were provided to us, the paintings

are significantly darker compared to their frame and do not correspond to the brightness of the paintings in the database. This leads to poor precision in the retrieval step. To cope with this problem, the brightness component is increased for each previously found bounding box.

Afterward, by transforming its format from BGR to HSV and then applying the otsu threshold again, we are able to obtain a more precise distinction between painting and frame/background, which will then be used during the retrieval and rectification steps.

III. PAINTING RETRIEVAL

For the retrieval of the paintings situated in the database we use an approach based on feature detection algorithms.

After consulting a paper that performs a comparative analysis between the most well known algorithms [1] to get a general idea of the strengths and weaknesses of the different methods available to us, we carried out some experiments focusing on SIFT, AKAZE and ORB.

The results obtained made us opt for ORB, because overall it gave us more precise results than AKAZE and, unlike SIFT, it's free of charge, therefore usable without fees in a possible commercial application.

To save time, the key points of the paintings in the database have been previously calculated and stored using the pickle module [link pickle documentation].

It implements binary protocols for serializing and de-serializing a Python object structure.

This data is loaded only once during the launch of the program. The key points are computed from the bounding boxes detected by the previously described pipeline and afterward, to determine the best matches, the ratio test proposed by D. Lowe in the SIFT paper is performed[Reference to the sift paper].

This measure is obtained by comparing the distance of the closest neighbor to that of the second-closest neighbor.

This measure performs well because correct matches need to have the closest neighbor significantly closer than the closest incorrect match to achieve reliable matching.

For false matches, there will likely be a number of other false

matches within similar distances due to the high dimensionality of the feature space.

In our implementation we reject all matches in which the distance ratio is greater than 0.75.

This allows us to keep the number of correctly retrieved paintings still high but at the same time to decrease the number of false positive. A ranking with the 5 best matches found is then created and saved to a CSV file to show the results.

To understand if the painting is actually recognized among those in the DB, the average of the key points matched among the best 5 ranked is calculated and, if the first one differs from it for more than significant value, it is considered as correct and shown on the interface.

IV. PAINTING RECTIFICATION

A. Four points transform

On the contour found with the techniques described in **Precision boosting and paintings segmentation** is applied the function `approxPolyDP` from OpenCV to approximate it to a polygonal curve.

If the shape returned has four vertices we can assume that the process has found a rectangular painting.

Given the four points, we are able to estimate the homography and apply the transformation to rectify the painting.

To calculate the aspect ratio for the projected rectangle we use an implementation based on this paper [] which derives the equations assuming a pinhole camera model.

This pipeline is applied to the paintings that don't get a match on the database.

B. alignImages

Not all the paintings inside the museum have a rectangular shape. This means that the approximation found by `approxPolyDP` does not consist of 4 vertices, making the method just explained impractical.

The approach that is used in this case is based on the common keypoints found between the distorted image and the corresponding match in the database.

Through ORB feature matching algorithm the key points are computed from each bounding boxes detected by the previous pipeline and subsequently, to determine the best matches between it and the images of the database, the ratio test is used.

From them we calculate the homography matrix and to avoid mismatches, the RANSAC algorithm is exploited.

To obtain a better result also the inverse warping algorithm is utilized.

V. PEOPLE DETECTION

A neural network (YoloV3) is used for people detection. Each video frame is passed through the network to make inference, and find all bounding boxes containing one of the objects in our classes list.

The weights for the network are obtained from an already trained network on COCO[reference to COCO paper], a famous dataset containing 80 different classes.

In our case the only wanted class is the person one, for ease of use the network has not been modified, but from its output will be deleted all the classes with an id different from 0 that is the id of the person class.

A little problem has arisen due to the high number of paintings representing persons, in fact the network detect them also inside the paintings, to prevent these false positive we added a new control on the pixel position in order to cut out all the people detected inside a bounding box previously classified as a painting.

VI. PEOPLE LOCALIZATION

People Localization

VII. PAINTING REPLACEMENT IN THE 3D MODEL

Painting Replacement in the 3D Model

A. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”.)

B. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

C. *LaTeX*-Specific Advice

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don't use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in L^AT_EX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you've discovered a new method of counting.

BIB_TE_X does not work by magic. It doesn't get the bibliographic data from thin air but from .bib files. If you use BIB_TE_X to produce a bibliography you must send the .bib files.

L^AT_EX can't read your mind. If you assign the same label to a subsubsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

L^AT_EX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it's supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

D. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.

- There is no period after the "et" in the Latin abbreviation "et al."
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [7].

E. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

F. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

G. Figures and Tables

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	Table Column Head		
	<i>Table column subhead</i>	<i>Subhead</i>	<i>Subhead</i>
copy	More table copy ^a		

^aSample of a Table footnote.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when



Fig. 1. Example of a figure caption.

writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M ”, not just “ M ”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

REFERENCES

- [1] Tareen, Shaharyar Ahmed Khan, and Zahra Saleem. “A comparative analysis of sift, surf, kaze, akaze, orb, and brisk.” 2018 International conference on computing, mathematics and engineering technologies (iCoMET). IEEE, 2018.
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] K. Elissa, “Title of paper if known,” unpublished.
- [5] R. Nicole, “Title of paper with only first word capitalized,” *J. Name Stand. Abbrev.*, in press.

- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [7] M. Young, *The Technical Writer’s Handbook*. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.