REAL TIME SYSTEMS IN COMPUTER VISION AND RECOGNITION OF IMAGES

ABSTRACT

The importance of micro-expression is that micro-expressions occur in a meaningless way, that is, they occur in a forced, disobedient way, thus revealing human emotions. These statements can be useful for criminal investigations, airport security, forensic science, and especially for psychological expertise. The article describes the development of models and algorithms for systematic analysis of digital camera data in real time.

KEYWORDS

Micro expression, real-time system, computer vision, 3D model, SVM, image, ip camera

In many practical areas, real-time is a peculiar concept. Here is an example from digital signal processing. If the processing of audio data requires more than T for each T second of analysis, such processing is not a real-time process. If less than T seconds is required, this is now a real time process. There is a “solid” and “soft” real-time system. The “solid” real-time system allows you to perform actions at a specific interval. Generally, such guarantees are required for systems that cause fatal impairment of the system's ability to exceed the established duration of the reaction.

The "soft" real-time system, as a rule, allows you to operate within a specified time interval. Exceeding the set time limit for such a system will result in reduced performance, but will not affect system performance. In most cases, software systems are "soft" real-time. Real-time software systems must be managed and run with a dedicated real-time operating system. These systems provide real-time support through special tools. These tools include:

Some operating systems (such as Linux) provide support for the aforementioned special tools for solving real-time tasks. However, they are strictly not a real-time operating system. When evaluating real-time systems, two important characteristics are used:

• interrupt response time - the time between the request to disconnect and the start of the interrupt processing function;

• control flow response time ("latency") - the time between requesting a disconnect and initiating an interrupt reaction flow. This includes time interruptions, scheduling pauses, and context changes.

In computer vision, researchers are developing mathematical techniques for comparing three-dimensional shapes and visions. There is now a reliable technical solution to accurately calculate the partial 3D model of the environment, from thousands of partially covered photographs (Figure 1.1 a). Given the large number of images of a particular object or face, we can create 3D dense 3D surface models using stereo matching (Figure 1.1b). We need to track the individual moving against a complex background (Figure 1.1 c). You might even try to photograph all the people in the photo using face recognition, clothing, and hair (Figure 1.1 d). However, in spite of all this, computer-generated imagery interprets the same age (for example, counting all animals in the picture). Why is the case so compelling? Partly because confrontation is the solution to the problem, and we try to recover some unknown people who have no clear information on the solution. Therefore, we need to turn to physics and probabilistic models to isolate potential solutions. However, modeling the visual world in all its complexity is more serious than modeling the vocal system that produces sound sounds [1].

Computer-based models are usually developed in physics (radiometry, optics and sensor design) and computer graphics. Both models model how the objects move and revive, how light on their surfaces is created, dispersed by the atmosphere, broken through camera lenses (or human eyes) and photographed primarily in the plane of the image. While computer graphics are not perfect (a fully computerized movie with human characters still managed to cross a spacious valley that separates them from real robots and computer-animated humans), in limited domains, such as everyday objects or dinosaurs are a virtual reality, like the revival of extinct creatures.

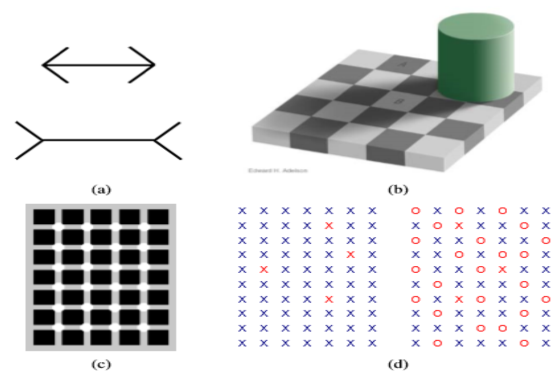

Figure 1.1. Examples of images used in computer vision

In computer viewing, we try to reverse the image we see in one or more of the images, the reconstruction features such as shape, light and color distribution. Doing so is easy for humans and animals, and for computer scanning algorithms, it is very difficult. However, today's computer vision is used in a variety of real-world applications:

• Optical character recognition (OCR): letters and automatic plate identification (ANPR) (Figure 1.2a);

• Machine inspection: fast-track quality control using specialized stereo imaging that meets airplane wings or body parts (Fig. 1.2b) or seeks to detect defects in X-ray textures;
• Retail: automatic automated detection (Figure 1.2c);
• 3D model building (photogrammetry): Automatically build 3D models of parts used in systems such as Bing Maps;
• Medical imaging: recording preoperative and intraoperative representations (Figure 1.2d) or conducting long-term studies on the morphology of the human brain during their lifetime;
• Vehicle Safety: Detection of unexpected pedestrians on the streets in the situations like active viewing methods such as radars are not working properly (Figure 1.2e);

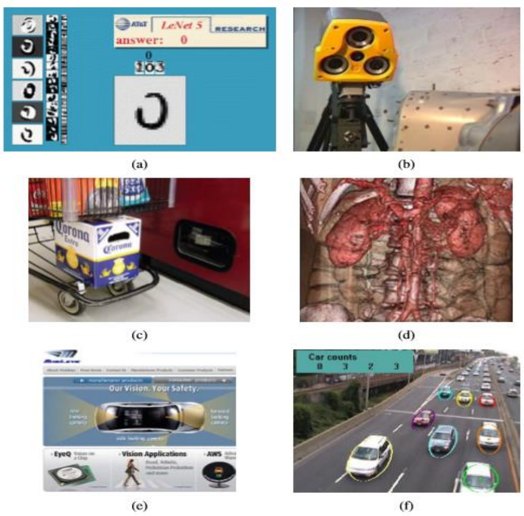

Figure 1.2. computer vision in various areas

• Meeting behavior: Combining computer-generated images (CGI) with live video and tracking feature points in the source video to evaluate 3D camera movement and environmental form. This technique is widely used in Hollywood; requires the use of new elements to insert between the front and back elements.
• Motion Capture (MoCap): using retro-rectangular characters from multiple cameras or other visual techniques to attract actors for computer animation;
• Observation: disturbance monitoring, road traffic analysis (Figure 1.2f);
• Fingerprint and biometrics detection: Automated authentication and forensic handbooks.

The following are some of the methods for detecting facial expression (micro-expression (ME) - micro expression) and suggest an effective method. The input method is assumed to be 25fps frame-rate video captured by a standard camera. The source of the video is in real-time, and does not need to work in advance in the laboratory under a precise scenario, but it moves freely. The camera is not fixed or stationary on permanent moving objects. Event coverage and detection may change. It is assumed that muscle contraction in ME (which may occur) results in a dramatic change in density within a given area of the face image. It can be affected in two ways: (1) texture change, e.g. wrinkles may occur, or (2) normal surface changes, if a large textured region is moved. Have both effects or one of them. Because of its small size and easily confused with the event's global head / camera movement, speech, blinking or simple (controlled) expression, a first-ever registration algorithm is proposed to abolish global movement. , a classification tool is then used to identify the common symptoms of ME from the recorded intensity difference signal [2].

The area under study is illustrated in Figure 1.3. The first face is found and its icons are identified sequentially in the video images. The facial image is then converted into a canonical coordinate system and is divided into investments that characterize one hundred parts of the actions that are triggered by the ROI (Region of interest). Each ROI measures the image density change in a temporary floating window. The SVM (Support Vector Machine) classifier has been trained to distinguish ME from the wrong density changes caused by other motions and changes in light.

Two methods for micro-expression detection are proposed. The first is that, without any experience, only high-resolution changes are in the apex frame of the ME. The other is controlled and used to obtain ME samples.

The main method

Let's take a look at Figure 1.4 and an example of a Φ_t compatible ME. The most sensitive ME frame (apex frame) should be the biggest difference between the front frames. Therefore, by combining the differences, one score is obtained

$$b(t) = \sum_{j=t-T}^{t-1} \left\| \hat{I}_t - \hat{I}_j \right\|_2^2$$

The designated signal b (t) can be subtracted by the sum of Φ_t .
The sign b(t) is paused. The maximum result is usually consistent with ME apex squares. The idea behind ME is that Φ_t the signals through the reference frames are consistent and cause a strong response b(t) as shown in Figure 1.5.

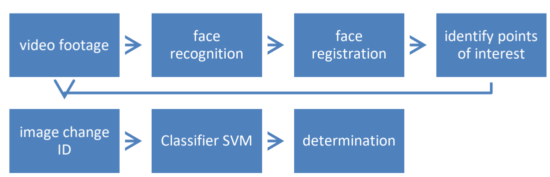

Figure 1.3. Flowchart of the proposed automatic ME detection algorithm.

However, shortcomings, other events that cause rapid change in other local maximal images, and so on, create global action or flashing. This vulnerability is partially mitigated by designing the classifier as described in the next section.

SVM classifier. The classifier is studied to differentiate the MES from other fluctuations in density. The SVM classifier with the RBF kernel is prepared in the SMIC_E_HS database [3]. In SMIC, videos have a 100fps frame rate. Therefore, sub-samples were selected by selecting the 4th element of each of the Φ_t vectors because our test videos are only 25 fps.

Positive samples were collected during the training phase. t_i is the initial frame of annotated MEs for each region, i = 1,… ..N. It is then identified as a positive specimen collection

$$P = \bigcup_{i=1}^{N} \{\Phi_{t_i-15}, \Phi_{t_i-14}, \dots, \Phi_{t_i+5}\}$$

Increasing the frame size of 15 squares and 5 squares after the first 100fps video starting frame ensures that the apex frame of ME is always present in the vectors. From the remaining video tapes negative samples were randomly collected in equal numbers [6]. It is noted that a specific SVM classifier teaches k = 1,… .., 12 for each region of interest. The reason for the independent processing is that the density model may be different due to points of interest and other textures. Finally, SVMs are taught in the study data from 800–2000 specimens depending on the occurrence of ME in a particular ROI [4].

System information. A few small tricks were made to improve the accuracy. The eye areas were removed from the visible detectors because the blinking of the eye was approximately the same instantaneous movement as the MES. However, eye blinking is unlikely, since accidental blinking is often partially consistent with MEs [5].

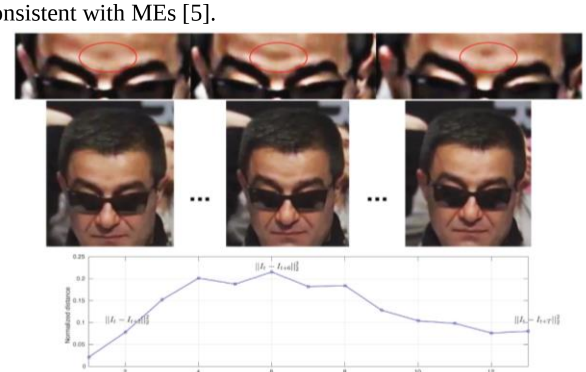

Figure 1.4. Micro Expression (Surprising Feeling) after Poker

Finally, the detector response was modified at maximum pressure with a 20-square window to avoid multiple responses for the same case. Two algorithms for facial features were checked - Face and Interface. Both methods come from the modern implementation of a cascade of linear regressors. One thing is that the Face is compatible with the 3D model and the Interface 2D. Detectors have shown that each video sequence contains a set of 49 facial features.

A few small tricks were made to improve the accuracy. The eye areas were removed from the visible detectors because the blinking of the eye was approximately the same instantaneous movement as the MES. However, the blinking of the eye cannot be avoided, since accidental blinking is often partially consistent with ME. In real-time videos, participants are not always recorded. Subsequently, a subset of ROIs are completely or partially closed due to the yaw.

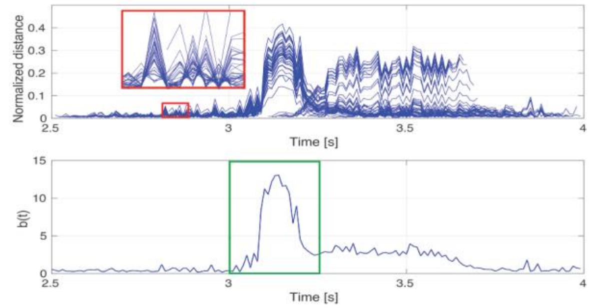

Figure 1.5. The upper plot shows Φ_t extended density frames through t reference frequencies. Signal consistency and the largest difference in the apex frame result in the peak of b (t)

Therefore, the previous position was determined, and the properties of the facial features were ignored in subsequent processing. The ratio of the inner eye angle and the highest angle of the nose was considered simple by calculating the natural logarithm. In the frontal position, the log-coefficient was approximately 0, but the log-coefficient of the frontal position was found to exceed the empirically fixed threshold [0: 5].

REFERENCES

[1] Musaev M.M. The development of spectral methods in the processing of signals and images. Bulletin of TUIT, No. 1, 2007 - p.14-18.

[2] Usmanov R.N., Kuchkarov T.A., Abdusalomov A.B. Computer vision. 2018 Tashkent, "Aloqachi".

[3] Bay H. SURF: Speeded Up Robust Features / H. Bay, A. Ess, T. Tuytelaars, Luc Van Gool // Computer Vision and Image Understanding, Vol. 110, No. 3, 2008. pp. 346-359

[4] Bennett J. LTE // LTE Encyclopedia. URL: https://sites.google.com/site/lteencyclopedia/home (Accessed: November 1, 2012)

[5] Chen Ting. Video Stabilization Algorithm Using a Block-Based Parametric Motion Model // Stanford University, Winter 2000, pp. 3-4. URL: http://scien.stanford.edu/pages/labsite/2000/ee392j/projects/chenreport.pdf (Date of access: 11.11.2012)

[6] Cohen P. OpenCL: What you need to know // MacWorld, August 2008. (Accessed: November 1, 2012)